

# LANOVA Penalization for Unreplicated Data

Maryclare Griffin<sup>1</sup> and Peter D. Hoff<sup>2</sup>

March 28, 2017

## Abstract

We consider the problem of estimating the entries of an unknown mean matrix or tensor,  $\mathbf{M}$ , given a single noisy realization,  $\mathbf{Y} = \mathbf{M} + \mathbf{Z}$ . In the matrix case, we address this problem by decomposing  $\mathbf{M}$  into a component that is additive in the rows and columns, i.e. the additive ANOVA decomposition of  $\mathbf{M}$ , plus a matrix of element-wise effects,  $\mathbf{C}$ , and assuming that  $\mathbf{C}$  may be sparse. Accordingly, we estimate  $\mathbf{M}$  by solving a penalized regression problem, applying a lasso penalty for elements of  $\mathbf{C}$ . We call the corresponding estimate of  $\mathbf{M}$  the LANOVA penalized estimate. Although solving this penalized regression problem is straightforward, specifying appropriate values of the penalty parameters is not. Leveraging the posterior mode interpretation of the penalized regression problem, we define and study moment-based empirical Bayes estimators of the penalty parameters. We show that our empirical Bayes estimators are consistent, and examine the behavior of LANOVA penalized estimates under misspecification of the distribution of elements of  $\mathbf{C}$ . We extend LANOVA penalization to accommodate sparsity of row and column effects and to tensor data. We demonstrate empirical Bayes LANOVA penalization in analyses of several datasets, including a matrix of microarray data, a three-way tensor of fMRI data and a three-way tensor of experimental data.

*Keywords:* adaptive estimation, method of moments, multiway data, structured data, transposable data, regularized regression.

<sup>1</sup>Department of Statistics, University of Washington-Seattle, Seattle, WA 98195 ([mgrffn@uw.edu](mailto:mgrffn@uw.edu)).

<sup>2</sup>Department of Statistical Science, Duke University, Durham, NC 27710 ([peter.hoff@duke.edu](mailto:peter.hoff@duke.edu)).

This research was supported by NSF grants DGE-1256082 and DMS-1505136. Replication code is available at <https://github.com/maryclare/LANOVA>.

# 1 Introduction

In many applications, researchers are interested in estimating the entries of an unknown  $n \times p$  mean matrix,  $\mathbf{M}$ , given a single noisy realization,  $\mathbf{Y} = \mathbf{M} + \mathbf{Z}$ , where the entries of  $\mathbf{Z}$  are assumed to be independent, identically distributed mean zero normal random variables with unknown variance  $\sigma_z^2$ . This estimation problem arises in various settings, from estimating a matrix of expression levels by gene and subject to estimating a tensor of treatment effects from a factorial experiment in the absence of replicate measurements.

This estimation problem is challenging: Each unknown parameter  $m_{ij}$  corresponds to a single observation,  $y_{ij}$ , and so the maximum likelihood estimate  $\mathbf{Y}$  has high variability. Accordingly, simplifying assumptions that reduce the dimensionality of  $\mathbf{M}$  are often made. Many common simplifying assumptions relate to the following representation of  $\mathbf{M}$ , based on a two-way ANOVA decomposition:

$$\mathbf{M} = \mu \mathbf{1}_n \mathbf{1}_p' + \mathbf{a} \mathbf{1}_p' + \mathbf{1}_n \mathbf{b}' + \mathbf{C}, \quad (1)$$

where  $\mu$  is an unknown grand mean,  $\mathbf{a}$  is an  $n \times 1$  vector of unknown row effects,  $\mathbf{b}$  is a  $p \times 1$  vector of unknown column effects,  $\mathbf{C}$  is a matrix of elementwise “interaction” effects and  $\mathbf{1}_n$  and  $\mathbf{1}_p$  are  $n \times 1$  and  $p \times 1$  vectors of ones, respectively. In the absence of replicates, many data analyses implicitly assume that  $\mathbf{C} = \mathbf{0}$ . This assumption reduces the number of freely varying unknown parameters substantially, from  $np$  to  $n + p$ , but it is also unlikely to hold in practice.

In the settings we consider, it is likely that at least some elements of  $\mathbf{C}$  are nonzero, e.g. some subjects-gene combinations or treatment pairs in a factorial design may have large interaction effects that are not explained by a strictly additive model. If  $\mathbf{M}$  is approximately additive in the sense that large deviations from additivity are rare, then  $\mathbf{C}$  is approximately sparse and estimation of  $\mathbf{M}$  may be improved by penalizing elements of  $\mathbf{C}$ :

$$\min_{\mu, \mathbf{a}, \mathbf{b}, \mathbf{C}} \frac{1}{2\sigma_z^2} \left\| \text{vec} \left( \mathbf{Y} - (\mu \mathbf{1}_n \mathbf{1}_p' + \mathbf{a} \mathbf{1}_p' + \mathbf{1}_n \mathbf{b}' + \mathbf{C}) \right) \right\|_2^2 + \lambda_c \left\| \text{vec}(\mathbf{C}) \right\|_1. \quad (2)$$

The  $\ell_1$  penalty induces sparsity among the estimated entries of  $\mathbf{C}$  and solving this penalized regression problem yields unique estimates of  $\mathbf{M}$  and  $\mathbf{C}$ .

The representation of  $\mathbf{C}$  is a departure from the existing literature, in which assumptions on the rank of  $\mathbf{M}$  or the rank of components of  $\mathbf{C}$  are made. Some set  $\mu = 0$ ,  $\mathbf{a} = \mathbf{0}$  and  $\mathbf{b} = \mathbf{0}$  and assume  $\mathbf{C}$  is low rank (Candes et al., 2013; Fazel, 2002; Gerard and Hoff, 2015; Hoff, 2007; Josse and Sardy, 2016), while others treat  $\mu$ ,  $\mathbf{a}$  and  $\mathbf{b}$  as unknown, apply the standard ANOVA zero-sum constraints for identifiability  $\mathbf{1}'_n \mathbf{a} = \mathbf{1}'_p \mathbf{b} = 0$  and  $\mathbf{C}' \mathbf{1}_n = \mathbf{0}$  and  $\mathbf{C} \mathbf{1}_p = \mathbf{0}$  and assume that  $\mathbf{C}$  is low rank (Gollob, 1968; Goodman and Haberman, 1990; Johnson and Graybill, 1972; Mandel, 1971; van Eeuwijk and Kroonenberg, 1998). All of these low rank methods correspond to a multiplicative model for elementwise effects. In the matrix case, they imply that elements of  $\mathbf{C}$  are multiplicative in row and column factors, i.e. if  $\mathbf{C}$  is rank  $R$  then  $c_{ij} = \sum_{r=1}^R \lambda_r u_{r,i} v_{r,j}$ .

Although clearly useful in many settings, these approaches have the two main limitations. First, in the presence of unknown noise variance,  $\sigma_z^2$ , existing methods for choosing the rank can be computationally expensive for large matrices (Hoff, 2007), require additional strong assumptions such as known  $\sigma_z^2$  (Candes et al., 2013), or rely on approximations to account for unknown  $\sigma_z^2$  that may not always perform well in practice (Josse and Sardy, 2016). Second, even when the rank can be chosen well, these methods conflate the presence of elementwise effects of scientific interest with the presence of multiplicative effects. While this may be plausible in many settings, it is easy to imagine scenarios in which a low rank multiplicative model may fail to capture elementwise effects of scientific interest, e.g. if  $\mathbf{Y}$  were an  $n \times n$  square matrix and all  $c_{ii}$  were large while all  $c_{ij}$ ,  $i \neq j$  were equal to zero, a rank  $n$  estimate of  $\mathbf{C}$  would be needed.

That being said, solving Equation (2) requires specification of two tuning parameters:  $\lambda_c$  and  $\sigma_z^2$ . Standard methods for specifying tuning parameters such as  $\lambda_c$  and  $\sigma_z^2$  are not well-suited to address this. Specifically, cross validation would require an intermediate imputation procedure that may slow computation while also requiring more assumptions that

may not be well justified. Alternatively, the  $\ell_1$  penalty on  $\mathbf{C}$  may be viewed as a Laplace prior distribution, in which case the standard approach for obtaining data-adaptive empirical Bayes estimates of  $\lambda_c$  and  $\sigma_z^2$  via maximum marginal likelihood may be prohibitively computationally demanding for high dimensional  $\mathbf{M}$  (Figueiredo, 2003; Kotz et al., 2001; Park and Casella, 2008).

In this paper we present an empirical Bayes approach to estimating  $\lambda_c$  and  $\sigma_z^2$  that is fast, easy to implement, consistent, and does not depend on assumptions made regarding lower-order mean parameters,  $\mathbf{a}$  and  $\mathbf{b}$ . We can then use these estimates of  $\lambda_c$  and  $\sigma_z^2$  to obtain additive-plus-sparse estimates of  $\mathbf{M}$  via Equation (2). As our approach to estimating  $\lambda_c$  and  $\sigma_z^2$  uses the Laplace prior distribution interpretation for the  $\ell_1$  penalty, we refer to estimation of  $\mathbf{M}$  via optimization of Equation (2) as Laplace ANOVA (LANOVA) penalization.

This paper proceeds as follows: In Section 2, we introduce simple moment-based empirical Bayes estimators of  $\lambda_c$  and  $\sigma_z^2$ , show that they are consistent as *either* the number of rows *or* columns goes to infinity, and show that their efficiency is comparable to the efficiency of asymptotically efficient marginal maximum likelihood estimators (MMLEs). In Section 3, we show how an estimate of  $\mathbf{M}$  can be obtained from Equation (2) given estimates of  $\lambda_c$  and  $\sigma_z^2$  using block coordinate descent, a standard method for solving Lasso regression problems. Additionally, we introduce a test of whether or not a sparse estimate of  $\mathbf{C}$  is appropriate, which is essentially a test of whether  $\mathbf{C}$  and  $\mathbf{Z}$  can be “deconvoluted,” and provide an investigation of robustness. In Section 4, we extend LANOVA penalization and the results of Sections 2 and 3 to the cases where we are interested in penalizing lower-order mean parameters  $\mathbf{a}$  and  $\mathbf{b}$  as well and also to the case where  $\mathbf{Y}$  is a  $K$ -way tensor. In Section 5, we apply LANOVA penalization model to a matrix of microarray data and two three-way tensors of experimental data collected using a three-way full factorial design. For the matrix of microarray data, we use LANOVA penalization as a preprocessing step to improve identification of genes associated with certain kinds of brain tumors. For one, we use an extension of LANOVA penalization that applies  $\ell_1$  penalties to lower order mean

parameters as well to perform an exploratory analysis of spatial patterns in fMRI activation response to different tasks. For the other, we use LANOVA penalization to identify “real” third order interaction effects in a three-way tensor of experimental data measuring infection of wheat varieties. Lastly, in Section 6, we discuss extensions of the concepts presented in this paper, specifically more general multilinear regression models and opportunities that arise in the presence of replicates.

## 2 LANOVA Tuning Parameter Estimation

Consider the following statistical model for deviations of an observed data matrix  $\mathbf{Y}$  from additivity:

$$\mathbf{Y} = \mathbf{A} + \mathbf{C} + \mathbf{Z} \tag{3}$$

$$\mathbf{A} = \mu \mathbf{1}_n \mathbf{1}_p' + \mathbf{a} \mathbf{1}_p' + \mathbf{1}_n \mathbf{b}'$$

$$\mathbf{C} = \{c_{ij}\} \sim \text{i.i.d. Laplace}(0, \lambda_c^{-1})$$

$$\mathbf{Z} = \{z_{ij}\} \sim \text{i.i.d. } N(0, \sigma_z^2).$$

The posterior mode of  $\mathbf{M} = \mathbf{A} + \mathbf{C}$  under this Laplace prior for  $\mathbf{C}$  and a flat prior for  $(\mu, \mathbf{a}, \mathbf{b})$  corresponds to the solution of LANOVA penalization problem given by Equation (2). The problem of specifying the tuning parameters for LANOVA penalization is equivalent to estimation of the unknown parameters  $\lambda_c$  and  $\sigma_z^2$ .

We construct estimators of  $\lambda_c$  and  $\sigma_z^2$  whose distribution is independent of the values of  $(\mu, \mathbf{a}, \mathbf{b})$  as follows. Letting  $\mathbf{H}_k = \mathbf{I}_k - \mathbf{1}_k \mathbf{1}_k' / k$  be the  $k \times k$  centering matrix, we define  $\mathbf{R}$  and show that  $\mathbf{R}$  depends on  $\mathbf{C}$  and  $\mathbf{Z}$  alone:

$$\begin{aligned} \mathbf{R} &= \mathbf{H}_n \mathbf{Y} \mathbf{H}_p \\ &= \mu \mathbf{H}_n \mathbf{1}_n \mathbf{1}_p' \mathbf{H}_p + \mathbf{H}_n \mathbf{a} \mathbf{1}_p' \mathbf{H}_p + \mathbf{H}_n \mathbf{1}_n \mathbf{b}' \mathbf{H}_p + \mathbf{H}_n (\mathbf{C} + \mathbf{Z}) \mathbf{H}_p \\ &= \mathbf{H}_n (\mathbf{C} + \mathbf{Z}) \mathbf{H}_p. \end{aligned}$$

We construct estimators of  $\lambda_c$  and  $\sigma_z^2$  from  $\mathbf{R}$  by leveraging the difference between Laplace and normal tail behavior, specifically, the difference in fourth order central moments. The fourth order central moment of any random variable  $x$  with mean  $\mu_x$  and variance  $\sigma_x^2$  can be expressed as  $\mathbb{E}[(x - \mu_x)^4] = (\kappa + 3)\sigma_x^4$ , where  $\kappa$  is interpreted as the *excess kurtosis* of the distribution of  $x$  relative to a normal distribution. A normally distributed variable has excess kurtosis equal to 0, whereas a Laplace distributed random variable has excess kurtosis equal to 3. It follows that the second and fourth order central moments of elements of  $\mathbf{C} + \mathbf{Z}$  are  $\mathbb{E}[(c_{ij} + z_{ij})^2] = \sigma_c^2 + \sigma_z^2$  and  $\mathbb{E}[(c_{ij} + z_{ij})^4] = 3\sigma_c^4 + 3(\sigma_c^2 + \sigma_z^2)^2$ , respectively, where  $\sigma_c^2 = 2/\lambda_c^2$ , the variance of a Laplace( $\lambda_c^{-1}$ ) random variable. Given values of  $\mathbb{E}[(c_{ij} + z_{ij})^2]$  and  $\mathbb{E}[(c_{ij} + z_{ij})^4]$ , we see that  $\sigma_c^2$  and  $\sigma_z^2$ , and accordingly  $\lambda_c$ , can easily be recovered. We do not observe  $\mathbf{C} + \mathbf{Z}$  directly, but we can use the the second and fourth order sample moments of  $\mathbf{R}$ , an estimate of  $\mathbf{C} + \mathbf{Z}$ , given by  $\bar{r}^{(2)} = \frac{1}{np} \sum_{i=1}^n \sum_{j=1}^p r_{ij}^2$  and  $\bar{r}^{(4)} = \frac{1}{np} \sum_{i=1}^n \sum_{j=1}^p r_{ij}^4$ , respectively, to separately consistently estimate  $\sigma_c^2$  and  $\sigma_z^2$  as both or either of  $n$  or  $p$  increase. These estimators are given as follows:

$$\begin{aligned} \hat{\sigma}_c^4 &= \left( \frac{n^3 p^3}{(n-1)(n^2 - 3n + 3)(p-1)(p^2 - 3p + 3)} \right) \left( \bar{r}^{(4)}/3 - (\bar{r}^{(2)})^2 \right) \\ \hat{\sigma}_c^2 &= \sqrt{\hat{\sigma}_c^4}, \quad \hat{\sigma}_z^2 = \left( \frac{np}{(n-1)(p-1)} \right) \bar{r}^{(2)} - \hat{\sigma}_c^2. \end{aligned} \quad (4)$$

An estimate of  $\lambda_c$  is then given by  $\hat{\lambda}_c = \sqrt{2/\hat{\sigma}_c^2}$ .

First, we examine the bias of  $\hat{\sigma}_c^4$ , which indirectly allows us to assess the finite sample performance of  $\hat{\sigma}_z^2$ ,  $\hat{\sigma}_c^2$  and  $\hat{\lambda}_c$ . Although it is possible to obtain an unbiased estimate of  $\sigma_c^4$ , the unbiased estimator will not be consistent as  $n \rightarrow \infty$  with  $p$  fixed or  $p \rightarrow \infty$  with  $n$  fixed. Accordingly, we prefer the biased estimator and examine its bias in the following proposition.

**Proposition 2.1** *Under the model given by Equation (3),*

$$\mathbb{E}[\hat{\sigma}_c^4] - \sigma_c^4 = - \left( \frac{n^3 p^3}{(n-1)(n^2 - 3n + 3)(p-1)(p^2 - 3p + 3)} \right) \left( \left( \frac{3(n-1)^2(p-1)^2}{n^3 p^3} \right) \sigma_c^4 + \left( \frac{2(n-1)(p-1)}{n^2 p^2} \right) (\sigma_c^2 + \sigma_z^2)^2 \right).$$

A proof of this proposition and all other results presented in this paper are given in an appendix. The bias is nonpositive for all values of  $n$  and  $p$ . When both  $n$  and  $p$  are relatively small,  $\hat{\sigma}_p^4$  tends to underestimate the magnitude of  $\mathbf{C}$  and overestimate the variability of elements of  $\mathbf{Z}$ . Recalling that  $\sigma_c^4$  is inversely related to  $\lambda_c$ , we can also interpret this result as stating that that we will tend to overpenalize and accordingly overshrink elements of  $\mathbf{C}$  when both  $n$  and  $p$  are small. This is not necessarily an undesirable result, in that it reflects a tendency to prefer the simple additive model when few data are available. We also observe that the bias of  $\hat{\sigma}_c^4$  depends not only on the variability of elements of  $\mathbf{C}$ , but also on the variability of elements of  $\mathbf{Z}$ . Holding  $n$ ,  $p$  and  $\sigma_c^2$  fixed, we will overestimate  $\lambda_c$  more when  $\sigma_z^2$  is larger. Again, this is not necessarily an undesirable result, in that it reflects a tendency to prefer the simple additive model when the data are very noisy. Lastly, we see that  $\mathbb{E}[\hat{\sigma}_c^4] - \sigma_c^4 = O\left(\frac{1}{np}\right)$ , i.e. the bias of  $\hat{\sigma}_c^4$  approaches zero as *either* the number of rows *or* the number of columns increases.

As we mentioned previously, our biased estimate given by Equation (4) is consistent as  $n \rightarrow \infty$  with  $p$  fixed,  $p \rightarrow \infty$  with  $n$  fixed or  $n, p \rightarrow \infty$ .

**Proposition 2.2** *Under the model given by Equation (3),  $\hat{\sigma}_c^4 \xrightarrow{p} \sigma_c^4$ ,  $\hat{\sigma}_c^2 \xrightarrow{p} \sigma_c^2$ ,  $\hat{\lambda}_c \xrightarrow{p} \lambda_c$  and  $\hat{\sigma}_z^2 \xrightarrow{p} \sigma_z^2$  as  $n \rightarrow \infty$  with  $p$  fixed,  $p \rightarrow \infty$  with  $n$  fixed, or  $n, p \rightarrow \infty$ .*

Because our empirical Bayes estimators are not maximum likelihood estimators, they may not be asymptotically efficient. We compare the efficiency of  $\hat{\sigma}_c^2$  to that of asymptotically efficient MML estimates, which require a more computationally demanding iterative algorithm to obtain, such as the EM algorithm (Park and Casella, 2008). The asymptotic variance of  $\hat{\sigma}_c^2$  is straightforward to compute as  $\sqrt{np}(\hat{\sigma}_c^4 - \sigma_c^4)$  converges in distribution to a

known moment estimator of  $\sigma_c^4$ . The asymptotic variance of the MMLE,  $\tilde{\sigma}_c^2$ , is given by the Cramér-Rao lower bound for  $\sigma_c^2$  which can be computed numerically from the known density of the sum of Laplace and normally distributed variables (Díaz-Francés and Montoya, 2008; Nadarajah, 2006). We plot the ratio,  $\mathbb{V}[\tilde{\sigma}_c^2]/\mathbb{V}[\hat{\sigma}_c^2]$  over values of  $\sigma_c^2, \sigma_z^2 \in [0, 1]$  in Figure 1. Note that the relative efficiency of our estimate of  $\hat{\sigma}_c^2$  compared to the MMLE of  $\tilde{\sigma}_c^2$  also reflects the relative efficiency of  $\hat{\lambda}_c$  and  $\hat{\sigma}_z^2$  compared to the MMLEs of  $\tilde{\lambda}_c$  and  $\tilde{\sigma}_c^2$ , respectively, because both are simple functions of  $\hat{\sigma}_c^2$ .

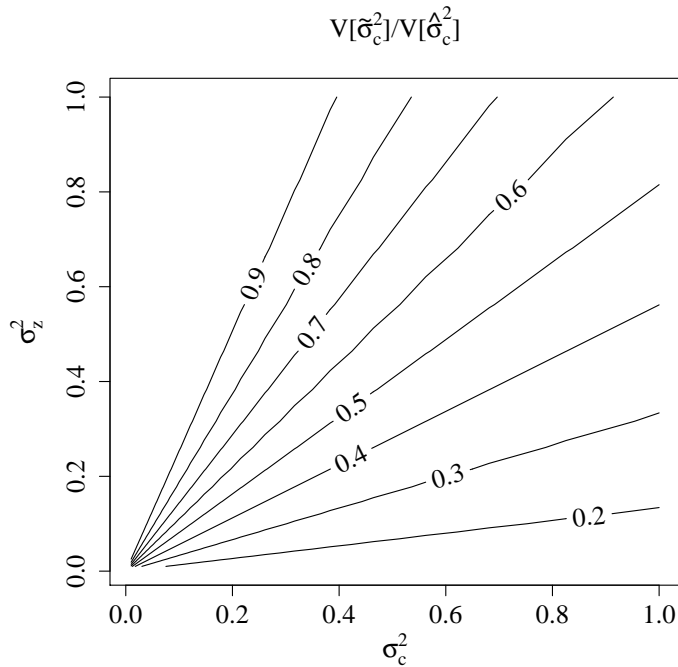


Figure 1: Relative asymptotic efficiency.

When  $\sigma_c^2$  is small relative to  $\sigma_z^2$ , the MMLE tends to be slightly more efficient but at the cost of a slower and more computationally demanding estimation procedure when  $\mathbf{M}$  is high dimensional. When  $\sigma_c^2$  is large relative to  $\sigma_z^2$ , the MMLE tends to be much more efficient, however the model will tend not to result in a simplified structure of the unknown matrix  $\mathbf{M}$  because interactions will not be heavily penalized. Put another way, Figure 1 indicates that  $\hat{\lambda}_c$  and  $\hat{\sigma}_z^2$  will have efficiency comparable to the MMLEs when the additive model explains much of the variability in  $\mathbf{M}$ , that is, when the interactions are sparse. We



also note that the empirical Bayes estimators are likely to be more robust to misspecification of the distribution of elements of  $\mathbf{C}$  than the MMLEs, insofar as the MMLE uses the entire likelihood to estimate  $\sigma_c^2$  and  $\sigma_z^2$  whereas our empirical Bayes estimators use the first through fourth moments.

### 3 Mean Estimation, Model Checking and Robustness

#### 3.1 Estimation

We can obtain estimates of  $\mathbf{M}$  via Equation (2) using any optimization method for Lasso regression. The results presented in this paper use block coordinate descent to obtain estimates of the unknown mean parameters by iterating the following procedure until the objective function given by Equation (2) converges, starting with  $\hat{\mathbf{C}}^{(0)} = \mathbf{H}_n \mathbf{Y} \mathbf{H}_p$  and  $k = 1$ :

- Set  $\hat{\mu}^{(k)} = \mathbf{1}_n' (\mathbf{Y} - \hat{\mathbf{C}}^{(k-1)}) \mathbf{1}_p / (np)$ ,  $\hat{\mathbf{a}}^{(k)} = \mathbf{H}_n' (\mathbf{Y} - \hat{\mathbf{C}}^{(k-1)}) \mathbf{1}_p / p$ ,  $\hat{\mathbf{b}}^{(k)} = \mathbf{H}_p (\mathbf{Y} - \hat{\mathbf{C}}^{(k-1)})' \mathbf{1}_n / n$  and  $\mathbf{R}^{(k)} = \mathbf{Y} - \hat{\mu}^{(k)} \mathbf{1}_n \mathbf{1}_p' - \hat{\mathbf{a}}^{(k)} \mathbf{1}_p' - \mathbf{1}_n (\hat{\mathbf{b}}^{(k)})'$ ;
- Set  $\hat{\mathbf{C}}^{(k)} = \text{sign}(\mathbf{R}^{(k)}) \left( \left| \mathbf{R}^{(k)} \right| - \hat{\lambda}_c \hat{\sigma}_z^2 \right)_+$ , where  $\text{sign}(\cdot)$  and the soft thresholding function,  $(\cdot)_+$ , are applied elementwise.

In practice, our empirical Bayes estimators are not guaranteed to be nonnegative and two special cases can arise. When  $\hat{\sigma}_c^4 < 0$ , we set  $\hat{\sigma}_c^2 = 0$  and  $\mathbf{C} = \mathbf{0}$  and obtain a strictly additive estimate of the unknown mean matrix,  $\hat{\mathbf{M}} = (\mathbf{I}_n - \mathbf{H}_n) \mathbf{Y} (\mathbf{I}_p - \mathbf{H}_p)$ . When  $\hat{\sigma}_z^2 < 0$ , we reset  $\hat{\sigma}_z^2 = 0$ , set  $\mathbf{C} = \mathbf{R}$  and obtain a strictly non-additive estimate of the unknown mean matrix,  $\hat{\mathbf{M}} = \mathbf{Y}$ .

Because LANOVA penalization resembles ANOVA decomposition, we briefly discuss interpretation of LANOVA penalized estimates. First, we note that the nonzero entries of  $\hat{\mathbf{C}}$  correspond to the  $r$  largest residuals from fitting a strictly additive model, where  $r$  is determined by  $\hat{\lambda}_c$  and  $\hat{\sigma}_z^2$ . Accordingly, we can interpret elements of  $\hat{\mathbf{C}}$  as evidence of nonadditivity. Additionally, because we do not constrain rows and columns of  $\hat{\mathbf{C}}$  to sum to zero, we cannot

interpret elements of  $\hat{\mathbf{C}}$  directly as population average interaction effects as in ANOVA decomposition, i.e.  $\hat{c}_{ij} \neq \mathbb{E}[y_{ij}] - \frac{1}{p} \sum_{j=1}^p \mathbb{E}[y_{ij}] - \frac{1}{n} \sum_{i=1}^n \mathbb{E}[y_{ij}] + \frac{1}{np} \sum_{i=1}^n \sum_{j=1}^p \mathbb{E}[y_{ij}]$ . For the same reason, interpretation of elements of  $\hat{\mathbf{A}} = \mathbf{1}_n \mathbf{1}_p' \hat{\mu} + \hat{\mathbf{a}} \mathbf{1}_p' + \mathbf{1}_n \hat{\mathbf{b}}'$  is not the same as the interpretation of the corresponding ANOVA estimates. To obtain estimates of interaction effects that have the standard ANOVA interpretation, we recommend performing a two-way ANOVA decomposition of  $\hat{\mathbf{M}}$ . In the appendix, we show that the lower-order contrasts obtained via ANOVA decomposition of  $\hat{\mathbf{M}}$  are identical to those obtained by performing an ANOVA decomposition of  $\mathbf{Y}$ , i.e. LANOVA penalization preserves lower-order ANOVA contrasts and changes only higher-order contrasts.

### 3.2 Testing

As mentioned in the Section 2, our estimates are obtained by assuming the distribution of entries of  $\mathbf{C}$  have tail behavior consistent with a Laplace distribution. It is natural to ask if this assumption is appropriate but difficult to devise a test of the appropriateness of the Laplace assumption specifically. Accordingly, we suggest a test of the more general assumption that elements of  $\mathbf{C}$  are heavy-tailed to check whether or not the assumption of heavy-tailed  $\mathbf{C}$  is justified by the data. When the distribution of elements of  $\mathbf{C}$  is heavy-tailed, the distribution of elements of  $\mathbf{C} + \mathbf{Z}$  will also be heavy-tailed and will have strictly positive excess kurtosis,  $\kappa > 0$ . In contrast, when elements of  $\mathbf{C}$  are either all zero or have a distribution with normal tails, elements of  $\mathbf{C} + \mathbf{Z}$  have excess kurtosis equal to exactly zero. We construct a test of the null hypothesis  $H_0: c_{ij} + z_{ij} \sim \text{i.i.d. } N(0, \sigma_c^2 + \sigma_z^2)$ , which encompasses the cases in which  $\mathbf{C} = \mathbf{0}$  or elements of  $\mathbf{C}$  are normally distributed. We can think of this as a test of deconvolvability of  $\mathbf{C} + \mathbf{Z}$ , where the null hypothesis is that deconvolution of  $\mathbf{C} + \mathbf{Z}$  is not possible.

**Proposition 3.1** *Assume  $\mathbf{Y} = \mu \mathbf{1}_n \mathbf{1}_p' + \mathbf{a} \mathbf{1}_p' + \mathbf{1}_n \mathbf{b}' + \mathbf{C} + \mathbf{Z}$ . As both  $n$  and  $p \rightarrow \infty$ , an asymptotically level- $\alpha$  test of  $H_0: c_{ij} + z_{ij} \sim \text{i.i.d. } N(0, \sigma_c^2 + \sigma_z^2)$  is obtained by rejecting  $H_0$*

when

$$\sqrt{np} \left( \frac{\hat{\sigma}_c^4}{\sqrt{\frac{8}{3}(\hat{\sigma}_c^2 + \hat{\sigma}_z^2)^2}} \right) > z_{1-\alpha},$$

where  $z_{1-\alpha}$  denotes the  $1 - \alpha$  quantile of the standard normal distribution.

This test, as opposed to other tests of normality, is useful in that it gives us power against the alternative where elements  $\mathbf{C}$  are heavy-tailed.

We compute the approximate power of this test under two specific heavy-tailed distributions for elements of  $\mathbf{C}$ : the mean zero Laplace distribution with variance  $\sigma_c^2$  assumed for LANOVA penalization and a Bernoulli-normal spike-and-slab distribution, where an element of  $\mathbf{C}$  is equal to exactly zero with probability  $1 - \pi_c$  and is normally distributed with variance  $\tau_c^2$  if not equal to exactly zero. Recall that the Laplace distribution has excess kurtosis  $\kappa = 3$ . In contrast, the Bernoulli-normal spike-and-slab distribution has excess kurtosis  $\kappa = 3(1 - \pi_c)/\pi_c$ .

**Proposition 3.2** *Assume that elements of  $\mathbf{C}$  are independent, identically distributed mean zero Laplace random variables with variance  $\sigma_c^2$ . Then as  $n$  and  $p \rightarrow \infty$ , the asymptotic power of the test described in Proposition 3.1 is:*

$$1 - \Phi \left( \frac{z_{1-\alpha} - \sqrt{\frac{3np}{8}} \left( \frac{\phi^4}{(\phi^2+1)^2} \right)}{\sqrt{1 + \left( \frac{68\phi^8 + 36\phi^6 + 9\phi^4}{(1+\phi^2)^4} \right)}} \right),$$

where  $\phi^2 = \sigma_c^2/\sigma_z^2$ .

Immediately, we can see that the power depends on the variance of the heavy-tailed effects,  $\sigma_c^2$ , and that of the Gaussian noise,  $\sigma_z^2$ , only through their ratio,  $\phi^2$ . Additionally, we can see that when  $\phi^2 = 0$ , i.e. when  $\sigma_c^2 = 0$  and all elements  $\mathbf{C}$  are exactly 0, the approximate power is equal to  $\alpha$ . This is because elements of  $\mathbf{C} + \mathbf{Z}$  are normally distributed when elements of  $\mathbf{C}$  are all exactly equal to zero. The power for  $\alpha = 0.05$ ,  $\phi^2 \in [0, 2]$  and  $np = \{100, 200, \dots, 1000\}$  is plotted in Figure 2. We observe that the power of the test is

increasing in  $\phi^2$ , and increasing more quickly when  $np$  is larger, i.e. when more data are available.

Now we consider the power under the spike-and-slab prior for elements of  $\mathbf{C}$ :

**Proposition 3.3** *Assume that elements of  $\mathbf{C}$  are independent, identically distributed mean zero Bernoulli-normal random variables, where an element of  $\mathbf{C}$  is equal to exactly zero with probability  $1 - \pi_c$ , and is normally distributed with mean zero and variance  $\tau_c^2$  otherwise.*

*Then as  $n$  and  $p \rightarrow \infty$ , the asymptotic power of the test described in Proposition 3.1 is:*

$$1 - \Phi \left( \frac{z_{1-\alpha} - \pi_c (1 - \pi_c) \left( \sqrt{\frac{3np}{8}} \left( \frac{\phi^4}{(\pi_c \phi^2 + 1)^2} \right) \right)}{\sqrt{1 + \pi_c (1 - \pi_c) \left( \frac{(20\pi_c^2 - 28\pi_c + 35)\phi^8 + 16(5 - \pi_c)\phi^6 + 72\phi^4}{8(\pi_c \phi^2 + 1)^4} \right)}}$$

where  $\phi^2 = \frac{\tau_c^2}{\sigma_z^2}$ .

Again, we can see that the approximate power depends on the variances of the nonzero effects,  $\tau_c^2$ , and the Gaussian noise,  $\sigma_z^2$ , only through their ratio,  $\phi^2$ . We also observe that when  $\pi_c = 0$ ,  $\pi_c = 1$  or  $\phi^2 = 0$ , i.e.  $\tau_c^2 = 0$ , the approximate power of the test is  $\alpha$ , which is consistent with the fact that elements of  $\mathbf{C} + \mathbf{Z}$  are normally distributed in these cases. The power for  $\alpha = 0.05$ ,  $\pi_c \in [0, 1]$ ,  $\phi^2 \in \{0, 0.2, \dots, 2\}$  and  $np = \{100, 1000\}$  is shown in Figure 2. We observe that the approximate power is always increasing in  $\phi^2$  and  $np$ . For fixed  $\phi^2$  and  $np$ , power diminishes as  $\pi_c$ , the probability of an element of  $\mathbf{C}$  being nonzero, approaches 0 or 1. The approximate power diminishes more rapidly as  $\pi_c$  approaches 1 than as  $\pi_c$  approaches 0, especially when  $\phi^2$  or  $np$  is small, which is what we would expect given that the distribution of the elements of  $\mathbf{C}$  becomes more normal as more elements of  $\mathbf{C}$  are nonzero.

Under the Laplace and Bernoulli-normal spike-and-slab distributions for elements of  $\mathbf{C}$ , we generally find that the test has high power when estimating  $\mathbf{C}$  separately from  $\mathbf{Z}$  might be particularly valuable, e.g. when elements of  $\mathbf{C}$  are large in magnitude relative to Gaussian noise and when many entries of  $\mathbf{C}$  are expected to be exactly zero.

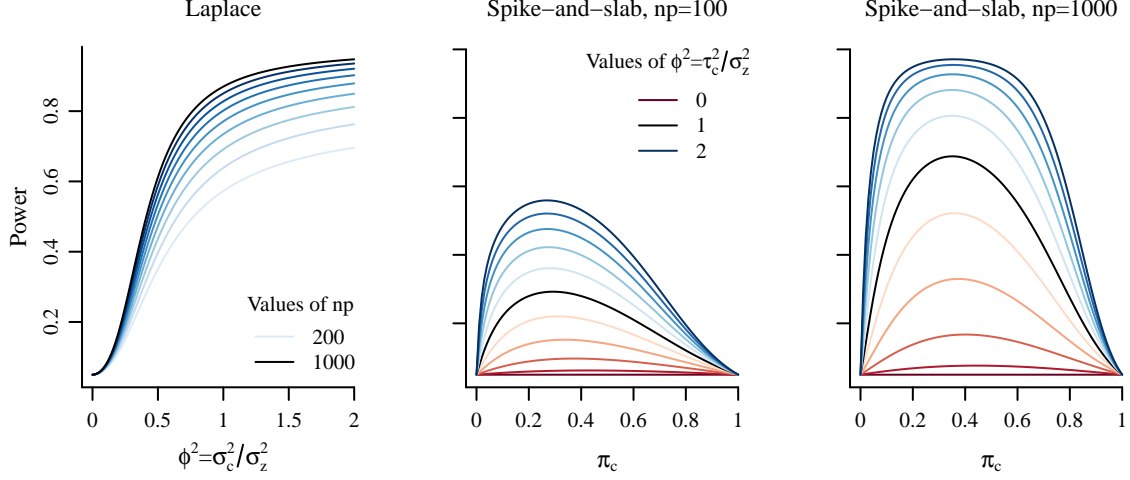


Figure 2: Approximate power of the test described in Proposition 3.1.

### 3.3 Robustness

Rejecting the null for the test described above suggests that the distribution of elements of  $\mathbf{C}$  is heavy-tailed, however it does *not* necessarily imply that the distribution of elements of  $\mathbf{C}$  is Laplace. If elements of  $\mathbf{C}$  are drawn from a different heavy-tailed distribution,  $\hat{\sigma}_c^2$  may not converge to the variance of  $\mathbf{C}$  and even if  $\hat{\sigma}_c^2$  does converge to the variance of  $\mathbf{C}$ , our posterior mode estimate of  $\mathbf{M}$  will not correspond to the mode under the “true” prior distribution. Naturally, this leads us to ask how our empirical Bayes estimates  $\hat{\sigma}_c^2$  and  $\hat{\sigma}_z^2$  and our estimate  $\hat{\mathbf{M}}$  perform when elements of  $\mathbf{C}$  are not drawn from a Laplace distribution. We find that the excess kurtosis,  $\kappa$ , of the “true” prior distribution for elements of  $\mathbf{C}$  determines our ability to estimate  $\sigma_c^2$  separately from  $\sigma_z^2$ .

**Proposition 3.4** *Under the model  $\mathbf{Y} = \mu \mathbf{1}_n \mathbf{1}_p' + \mathbf{a} \mathbf{1}_p' + \mathbf{1}_n \mathbf{b}' + \mathbf{C} + \mathbf{Z}$ , where elements of  $\mathbf{C}$  are independent, identically distributed draws from a mean zero, symmetric distribution with variance  $\sigma_c^2$ , excess kurtosis  $\kappa$  and finite eighth moment and elements of  $\mathbf{Z}$  are normally distributed with mean zero and variance  $\sigma_z^2$ ,  $\hat{\sigma}_c^2 \xrightarrow{p} \sqrt{\kappa/3} \sigma_c^2$  and  $\hat{\sigma}_z^2 \xrightarrow{p} \sigma_z^2 + \left(1 - \sqrt{\kappa/3}\right) \sigma_c^2$  as  $n \rightarrow \infty$  with  $p$  fixed,  $p \rightarrow \infty$  with  $n$  fixed, or  $n$  and  $p \rightarrow \infty$ .*

Proposition 3.4 indicates that we underestimate  $\sigma_c^2$  and overestimate  $\sigma_z^2$  when elements of  $\mathbf{C}$  are lighter tailed than LANOVA penalization assumes, and we overestimate  $\sigma_c^2$  and

underestimate  $\sigma_z^2$  when elements of  $\mathbf{C}$  are heavier tailed than LANOVA penalization assumes.

To see how this affects estimation of  $\mathbf{M}$ , we compare the risk of  $\hat{\mathbf{M}}$ , our estimate of  $\mathbf{M}$  under LANOVA penalization, to the risk of the maximum likelihood estimate (MLE) of  $\mathbf{M}$ ,  $\mathbf{Y}$  and the risk of the strictly additive estimate of  $\mathbf{M}$  obtained by solving the unpenalized problem with  $\mathbf{C} = \mathbf{0}$ . In general, the risk of  $\hat{\mathbf{M}}$  is not available in closed form. We focus on the Bernoulli-normal spike-and-slab distribution for elements of  $\mathbf{C}$  and compute Monte Carlo estimates of the relative risks, setting  $n = p = 25$ ,  $\mu = 0$ ,  $\mathbf{a} = \mathbf{0}$ ,  $\mathbf{b} = \mathbf{0}$ ,  $\sigma_z^2 = 1$ , and varying  $\tau_c^2 = \{\frac{1}{2}, 1, 2\}$  and  $\pi_c = \{0, 0.1, \dots, 0.9, 1\}$ . For each value of  $\tau_c^2$ , the variance of nonzero elements of  $\mathbf{C}$ , and  $\pi_c$ , the probability any element of  $\mathbf{C}$  is nonzero, the Monte Carlo estimate is based on 500 simulated values of  $\mathbf{Y} = \mathbf{C}$ .

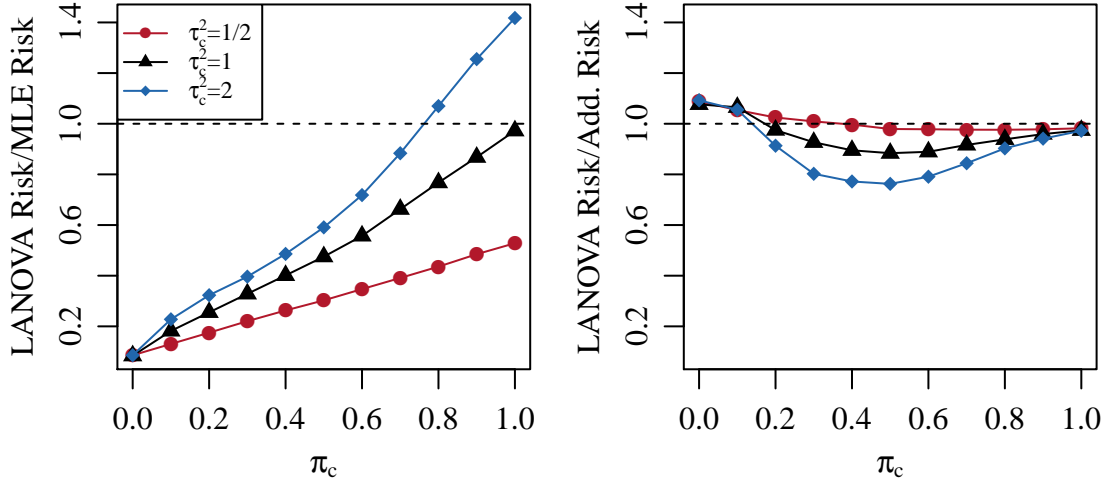


Figure 3: Monte Carlo approximations of relative risk of LANOVA estimate  $\hat{\mathbf{M}}$  versus the MLE,  $\mathbf{Y}$  and the strictly additive estimate of  $\mathbf{M}$ .

We observe that the LANOVA penalized estimate always outperforms the MLE when  $\tau_c^2 \leq \sigma_z^2$ . When  $\tau_c^2 > \sigma_z^2$ , we observe that the LANOVA penalized estimate still outperforms the MLE when  $\pi_c$  is relatively small, i.e. when  $\mathbf{C}$  is relatively sparse. This suggests that when we believe that the variability of elements of  $\mathbf{Z}$  exceeds the variability of nonzero elements of  $\mathbf{C}$  and/or that  $\mathbf{C}$  is very sparse, the LANOVA penalized estimate is likely to provide an improvement over the MLE. Additionally, we find that the LANOVA penalized

estimate outperforms the strictly additive estimate when nonnegligible nonadditive structure is present, especially when the variability of the elements of  $\mathbf{C}$  is large relative to the variability of elements of  $\mathbf{Z}$ .

In Figure 3, we also observe that the LANOVA penalized estimate performs well relative to both the MLE and the strictly additive estimate when  $\pi_c \approx 0.5$  and  $\kappa \approx 3$ , as assumed by LANOVA penalization. This suggests that cases of poor performance of the LANOVA estimate can be attributed to misspecification of  $\kappa$ . Accordingly, a simple correction of multiplying  $\hat{\sigma}_c^2$  by  $\sqrt{3/\kappa}$  if an alternative, more appropriate value of  $\kappa$  is known could result in improved performance of the LANOVA penalized estimate. In general, a more appropriate value of  $\kappa$  may not be known *a priori*, as excess kurtosis is not a readily interpretable quantity. However, when prior information suggests a Bernoulli-normal distribution is appropriate for modeling elements of  $\mathbf{C}$ , Proposition 3.4 suggests a correction for our empirical Bayes estimators that is based only on the expected number of nonzero elements of  $\mathbf{C}$ . Recalling that the excess kurtosis of the Bernoulli-normal spike-and-slab distribution is  $\kappa = 3(1 - \pi_c)/\pi_c$ , we can correct  $\hat{\sigma}_c^2$  by dividing  $\hat{\sigma}_c^2$  by  $\sqrt{\pi_c(1 - \pi_c)}$ . If we are right about  $\pi_c$  being close to the “true” proportion of nonzero elements of  $\mathbf{C}$ , then this will improve estimation of  $\sigma_c^2$  and  $\sigma_z^2$ .

We note that a correction that improves estimation of  $\sigma_c^2$  and  $\sigma_z^2$  may still not yield optimal estimation of  $\mathbf{M}$ , insofar as  $\hat{\mathbf{M}}$  computed from corrected estimates of  $\sigma_c^2$  and  $\sigma_z^2$  will not correspond to the posterior mode of  $\mathbf{M}$  under the distribution the corrected value of  $\kappa$  corresponds to. However given that obtaining the true posterior mode of  $\mathbf{M}$  under a spike-and-slab prior distribution is typically computationally infeasible, the corrected LANOVA penalized estimate may perform well relative to feasible alternatives, e.g. the MLE,  $\mathbf{Y}$ , or the strictly additive model.

## 4 Extensions

In this section, we describe two extensions to LANOVA penalization. The first extension involves penalizing lower-order parameters,  $\mathbf{a}$  and  $\mathbf{b}$ , in addition to the higher order mean parameters,  $\mathbf{C}$ . The second involves tensor-variate  $\mathbf{Y}$ .

### 4.1 Penalizing Lower-Order Parameters

When the matrix  $\mathbf{Y}$  has many rows and/or columns, setting  $\mathbf{C} = \mathbf{0}$  may still yield a model with many mean parameters. Furthermore, there may be scientific reasons to believe that many elements of  $\mathbf{a}$  or elements of  $\mathbf{b}$  are exactly zero. A natural extension of Equation (2) is given by

$$\min_{\mu, \mathbf{a}, \mathbf{b}, \mathbf{C}} \frac{1}{2\sigma_z^2} \|\text{vec}(\mathbf{Y} - \mathbf{M})\|_2^2 + \lambda_a \|\mathbf{a}\|_1 + \lambda_b \|\mathbf{b}\|_1 + \lambda_c \|\text{vec}(\mathbf{C})\|_1, \quad (5)$$

where we still have  $\mathbf{M} = \mathbf{1}_n \mathbf{1}_p' \mu + \mathbf{a} \mathbf{1}_p' + \mathbf{1}_n \mathbf{b}' + \mathbf{C}$ . Again, using the posterior mode interpretation of Equation (5), we can estimate  $\sigma_a^2$  and  $\sigma_b^2$  from the observed data,  $\mathbf{Y}$ :

$$\hat{\sigma}_a^2 = \frac{1}{n-1} \sum_{i=1}^n \check{a}_i^2 - \frac{n}{(n-1)(p-1)} \bar{r}^{(2)}, \quad \hat{\sigma}_b^2 = \frac{1}{p-1} \sum_{j=1}^p \check{b}_j^2 - \frac{p}{(n-1)(p-1)} \bar{r}^{(2)}.$$

where  $\check{\mathbf{a}} = \mathbf{H}_n \mathbf{Y} \mathbf{1}_p / p$  and  $\check{\mathbf{b}} = \mathbf{H}_p \mathbf{Y}' \mathbf{1}_n / n$  are OLS estimates for  $\mathbf{a}$  and  $\mathbf{b}$  for the unpenalized problem. The estimators  $\hat{\lambda}_a = \sqrt{2/\hat{\sigma}_a^2}$  and  $\hat{\lambda}_b = \sqrt{2/\hat{\sigma}_b^2}$  can be shown to be consistent for  $\lambda_a$  and  $\lambda_b$  as  $n \rightarrow \infty$  and  $p \rightarrow \infty$ , respectively. Because  $\hat{\lambda}_c$  and  $\hat{\sigma}_z^2$  are computed from functions of the data that are independent of  $\mathbf{a}$  and  $\mathbf{b}$ , our estimators for  $\lambda_c$  and  $\sigma_z^2$  are unchanged.

Given estimates of  $\lambda_a$ ,  $\lambda_b$ ,  $\lambda_c$  and  $\sigma_z^2$ , Equation (5) is still a standard Lasso regression problem and can be solved using any optimization method for Lasso regression. Again, we use block coordinate descent to obtain estimates of the unknown mean parameters by iterating the following procedure until the objective function given by Equation (5) converges, starting with  $\hat{\mathbf{a}}^{(0)} = \mathbf{H}_n \mathbf{Y} \mathbf{1}_p / p$ ,  $\hat{\mathbf{b}}^{(0)} = \mathbf{H}_p \mathbf{Y}' \mathbf{1}_n / n$  and  $\hat{\mathbf{C}}^{(0)} = \mathbf{H}_n \mathbf{Y} \mathbf{H}_p$  and  $k = 1$ :

- Set  $\hat{\mu}^{(k)} = \mathbf{1}_n' \left( \mathbf{Y} - \hat{\mathbf{a}}^{(k-1)} \mathbf{1}_p' - \mathbf{1}_n \left( \hat{\mathbf{b}}^{(k-1)} \right)' - \hat{\mathbf{C}}^{(k-1)} \right) \mathbf{1}_p / (np)$ ;



- Set  $\mathbf{R}_a^{(k)} = \mathbf{Y} - \hat{\mu}^{(k)} \mathbf{1}_n \mathbf{1}_p' - \mathbf{1}_n \left( \hat{\mathbf{b}}^{(k-1)} \right)' - \hat{\mathbf{C}}^{(k-1)}$  and set  $\hat{\mathbf{a}}^{(k)} = \text{sign} \left( \mathbf{H}_n \mathbf{R}_a^{(k)} \mathbf{1}_p / p \right) \left( \left| \mathbf{H}_n \mathbf{R}_a^{(k)} \mathbf{1}_p / p \right| - \hat{\lambda}_a \hat{\sigma}_z^2 / p \right)_+;$
- Set  $\mathbf{R}_b^{(k)} = \mathbf{Y} - \hat{\mu}^{(k)} \mathbf{1}_n \mathbf{1}_p' - \hat{\mathbf{a}}^{(k)} \mathbf{1}_p' - \hat{\mathbf{C}}^{(k-1)}$  and set  $\hat{\mathbf{b}}^{(k)} = \text{sign} \left( \mathbf{H}_p \left( \mathbf{R}_b^{(k)} \right)' \mathbf{1}_n / n \right) \left( \left| \mathbf{H}_p \left( \mathbf{R}_b^{(k)} \right)' \mathbf{1}_n / n \right| - \hat{\lambda}_b \hat{\sigma}_z^2 / n \right)_+;$
- Set  $\mathbf{R}_c^{(k)} = \mathbf{Y} - \hat{\mu}^{(k)} \mathbf{1}_n \mathbf{1}_p' - \hat{\mathbf{a}}^{(k)} \mathbf{1}_p' - \mathbf{1}_n \hat{\mathbf{b}}^{(k)}$  and set  $\hat{\mathbf{C}}^{(k)} = \text{sign} \left( \mathbf{R}_c^{(k)} \right) \left( \left| \mathbf{R}_c^{(k)} \right| - \hat{\lambda}_c \hat{\sigma}_z^2 \right)_+.$

We note that the lower-order contrasts obtained via ANOVA decomposition of  $\hat{\mathbf{M}}$  when lower-order mean parameters are penalized as well will no longer be identical to those obtained by performing an ANOVA decomposition of  $\mathbf{Y}$ .

## 4.2 Tensor Data

The results of Sections 2 and 3 for matrix-variate  $\mathbf{Y}$  can be extended to tensor data, where  $\mathbf{Y}$  is a  $K$ -way tensor. When  $\mathbf{Y}$  is a  $K$ -way tensor with dimensions  $p_1 \times p_2 \times \cdots \times p_K$ , LANOVA penalization is given by:

$$\text{vec}(\mathbf{Y}) = \text{vec}(\mathbf{A}) + \text{vec}(\mathbf{C}) + \text{vec}(\mathbf{Z}) \quad (6)$$

$$\text{vec}(\mathbf{A}) = \mathbf{W}\boldsymbol{\beta}$$

$$\mathbf{C} = \{c_{i_1 \dots i_K}\} \sim \text{i.i.d. Laplace}(0, \lambda_c^{-1})$$

$$\mathbf{Z} = \{z_{i_1 \dots i_K}\} \sim \text{i.i.d. N}(0, \sigma_z^2)$$

where  $\text{vec}(\mathbf{Y})$  is the  $\prod_{k=1}^K p_k$  vectorization of the  $K$ -mode tensor  $\mathbf{Y}$  with “lower” indices moving “faster”,  $\mathbf{W}$  is the design matrix corresponding to a  $K$ -way ANOVA decomposition treating the  $K$  modes of  $\mathbf{Y}$  as factors and  $\boldsymbol{\beta}$  is the vector of  $K$ -way ANOVA decomposition mean parameters. The matrix  $\mathbf{W} = [\mathbf{W}_1, \dots, \mathbf{W}_{2^K-1}]$  is obtained by concatenating the  $2^K - 1$  unique matrices of the form

$$\mathbf{W}_l = (\mathbf{W}_{l,1} \otimes \cdots \otimes \mathbf{W}_{l,K}),$$

where each  $\mathbf{W}_{l,k}$  is equal to either  $\mathbf{I}_{p_k}$  or  $\mathbf{1}_{p_k}$ , excluding the identity matrix,  $\mathbf{I}_{p_K} \otimes \cdots \otimes \mathbf{I}_{p_1}$ . In the three-way tensor case, the first part of Equation (6) refers to the following decomposition:

$$y_{ijk} = \mu + a_i + b_j + d_k + e_{ij} + f_{ik} + g_{jk} + c_{ijk} + z_{ijk},$$

for  $i = 1, \dots, p_1$ ,  $j = 1, \dots, p_2$  and  $k = 1, \dots, p_3$ . As in the matrix case, we initially focus on penalizing elements of the highest order mean term,  $\mathbf{C}$ , for which no replicates are observed.

We obtain estimates of  $\sigma_z^2$  and  $\lambda_c$  as follows. Our estimates are constructed from  $\text{vec}(\mathbf{R}) = (\mathbf{H}_K \otimes \cdots \otimes \mathbf{H}_1) \text{vec}(\mathbf{Y})$ , where  $\mathbf{H}_k = \mathbf{I}_{p_k} - \mathbf{1}_{p_k} \mathbf{1}_{p_k}^T / p_k$  is the  $p_k \times p_k$  centering matrix. The operation of multiplying  $\text{vec}(\mathbf{Y})$  by the matrix  $(\mathbf{H}_K \otimes \cdots \otimes \mathbf{H}_1)$  can be thought of as a sequence of  $K$  very simple operations. First, we rearrange  $\mathbf{Y}$  into a  $p_1 \times \prod_{k=2}^K p_k$  matrix, where the rows correspond to different levels of the first mode, and subtract off the row means. We then take the new matrix with row means subtracted off and rearrange it into a  $p_2 \times p_1 \prod_{k=3}^K p_k$  matrix, where the rows correspond to different levels of the second mode, and subtract off the row means. We repeat this procedure, taking the new matrix, rearranging it so that the rows correspond to levels of the  $k$ -th mode and subtracting off the row means for the remaining modes.

As in the matrix case,  $\text{vec}(\mathbf{R})$  not only contains information about both  $\lambda_c$  and  $\sigma_z^2$  but also is independent of the lower-order unknown mean parameters,  $\boldsymbol{\beta}$ , i.e.  $\text{vec}(\mathbf{R}) = (\mathbf{H}_K \otimes \cdots \otimes \mathbf{H}_1) \text{vec}(\mathbf{C} + \mathbf{Z})$ . Our estimates of  $\sigma_z^2$  and  $\lambda_c$  are still functions of the second and fourth sample moments of  $\mathbf{R}$ :  $\bar{r}^{(2)} = \frac{1}{p} \sum_{i=1}^p r_i^2$  and  $\bar{r}^{(4)} = \frac{1}{p} \sum_{i=1}^p r_i^4$ , where  $p = \prod_{k=1}^K p_k$ . We extend our empirical Bayes estimators as follows:

$$\begin{aligned} \hat{\sigma}_c^4 &= \left( \prod_{k=1}^K \frac{p_k^3}{(p_k - 1)(p_k^2 - 3p_k + 3)} \right) \left( \bar{r}^{(4)} / 3 - (\bar{r}^{(2)})^2 \right) \\ \hat{\sigma}_c^2 &= \sqrt{\hat{\sigma}_c^4}, \quad \hat{\sigma}_z^2 = \left( \prod_{k=1}^K \frac{p_k}{p_k - 1} \right) \bar{r}^{(2)} - \hat{\sigma}_c^2, \end{aligned} \tag{7}$$

where  $\hat{\lambda}_c = \sqrt{2/\hat{\sigma}_c^2}$ .

As in the matrix case, we can compute the bias of  $\hat{\sigma}_c^4$ :

**Proposition 4.1** *Under the model given by Equation (6),*

$$\mathbb{E} [\hat{\sigma}_c^4] - \sigma_c^4 = - \left( \prod_{k=1}^K \frac{p_k^3}{(p_k - 1)(p_k^2 - 3p_k + 3)} \right) \left( \left( 3 \prod_{k=1}^K \frac{(p_k - 1)^2}{p_k^3} \right) \sigma_c^4 + \left( 2 \prod_{k=1}^K \frac{p_k - 1}{p_k^2} \right) (\sigma_c^2 + \sigma_z^2)^2 \right).$$

The interpretation of this result is analogous to the matrix case; the bias indicates we tend to prefer the simpler model with  $\text{vec}(\mathbf{C}) = \mathbf{0}$  over a more complicated model with nonzero elements of  $\text{vec}(\mathbf{C})$  when few data are available and/or when the data is very noisy. Additionally,  $\mathbb{E} [\hat{\sigma}_c^4] - \sigma_c^4 = O\left(\frac{1}{p}\right)$ , i.e. the bias of  $\hat{\sigma}_c^4$  approaches zero as the number of levels of *any* mode increases.

We also assess the large-sample performance of our empirical Bayes estimators in the  $K$ -way tensor case.

**Proposition 4.2** *Under the model given by Equation (6),  $\hat{\sigma}_c^4 \xrightarrow{p} \sigma_c^4$ ,  $\hat{\sigma}_c^2 \xrightarrow{p} \sigma_c^2$ ,  $\hat{\lambda}_c \xrightarrow{p} \lambda_c$  and  $\hat{\sigma}_z^2 \xrightarrow{p} \sigma_z^2$  as  $p_{k'} \rightarrow \infty$  with  $p_k, k \neq k'$ , fixed or  $p_1, \dots, p_K \rightarrow \infty$ .*

The results for testing the appropriateness of assuming heavy-tailed  $\mathbf{C}$  and robustness also carry over to  $K$ -way tensors. Analogous results to Propositions 3.1-3.4, where we replace  $np$  with  $p$  and assume all  $p_1, \dots, p_K \rightarrow \infty$  can be shown to hold. We can also easily extend the block coordinate descent algorithm we used to estimate the unknown mean parameters in the matrix case:

- Set  $\hat{\boldsymbol{\beta}}^{(k)} = \text{argmin}_{\boldsymbol{\beta}} \left\| \text{vec}(\mathbf{Y}) - \text{vec}(\hat{\mathbf{C}}^{(k-1)}) - \mathbf{X}\boldsymbol{\beta} \right\|_2^2$ ;
- Set  $\text{vec}(\hat{\mathbf{C}}^{(k)}) = \text{sign} \left( \text{vec}(\mathbf{Y}) - \mathbf{X}\hat{\boldsymbol{\beta}}^{(k)} \right) \left( \left| \text{vec}(\mathbf{Y}) - \mathbf{X}\hat{\boldsymbol{\beta}}^{(k)} \right| - \hat{\lambda}_c \hat{\sigma}_z^2 \right)_+$ .

The step of setting  $\hat{\boldsymbol{\beta}}^{(k)}$  is simpler than it appears, because as we have observed in the matrix case, the unpenalized regression problem is the equivalent to fitting a  $K$ -way ANOVA decomposition to  $\mathbf{Y} - \hat{\mathbf{C}}^{(k-1)}$ . Lastly, by the same logic of the previous section, we can also extend LANOVA penalization for tensor data to penalize lower-order mean parameters as

well. When we observe tensor-variate  $\mathbf{Y}$ , penalizing lower-order parameters can be particularly useful because the “simplified” model with  $\mathbf{C} = \mathbf{0}$  still contains many parameters and is likely to be difficult to interpret. We give tuning parameter estimators for penalizing lower-order parameters in the three-way case in the appendix.

## 5 Numerical Examples

We illustrate the use of LANOVA penalization for three real data sets: a matrix of microarray data measuring gene expression levels, a a three-way tensor of experimental rat growth data and a three-way tensor of experimental data measuring infection of wheat varieties.

**Tumor Data:** In this example, we are interested in improving identification of genes that are differentially expressed in glioblastomas relative to oligodendroglial tumors by using LANOVA penalization to denoise the microarray data before performing principal components analysis. Glioblastomas differ from oligodendroglial tumors in their invasiveness. Glioblastomas are the most invasive, while oligodendroglial tumors include several less invasive gliomas: astrocytomas, oliodendroglomas, and mixed oligoastrocytomas. de Tayrac et al. (2009) used data on 24 glioblastomas and 19 oligodendroglial tumors to identify genes that are differentially expressed in glioblastomas versus oligodendroglial tumors using multiple factor analysis. We consider the same question and use a subset of the data made publicly available as part of the `denoiseR` package for R, which includes gene expression levels all 43 brain tumors and 356 continuous gene expression measurements (Josse et al., 2016). This data can be represented as a  $43 \times 356$  matrix.

We expect that the majority of the genes measured are not differentially expressed, i.e., many gene-by-tumor interaction effects are likely to be exactly zero. Accordingly, we apply LANOVA penalization with penalized interaction effects and unpenalized main effects to the matrix of microarray data before performing principal components analysis. The test given by Proposition 3.1 supports the use of a nonadditive estimate of  $\mathbf{M}$ ; we obtain a test statistic

of 18.45 and reject the null hypothesis of normally distributed elementwise variability at level  $\alpha = 0.05$ . We estimate that 11,188 elements of  $\mathbf{C}$  (73%) are exactly equal to zero.

Now, we perform principal components analysis on  $\hat{\mathbf{M}}$  and examine the first two principal components,  $\tilde{\mathbf{u}}_1$  and  $\tilde{\mathbf{u}}_2$ . Figure 4 shows that the  $\tilde{\mathbf{u}}_2$  separates glioblastomas and oligodendroglial tumors very well. This improves on the separation given by the second principal component of the raw data,  $\mathbf{Y}$ , shown in the second panel as well as the separation given by multiple factor analysis shown in de Tayrac et al. (2009).

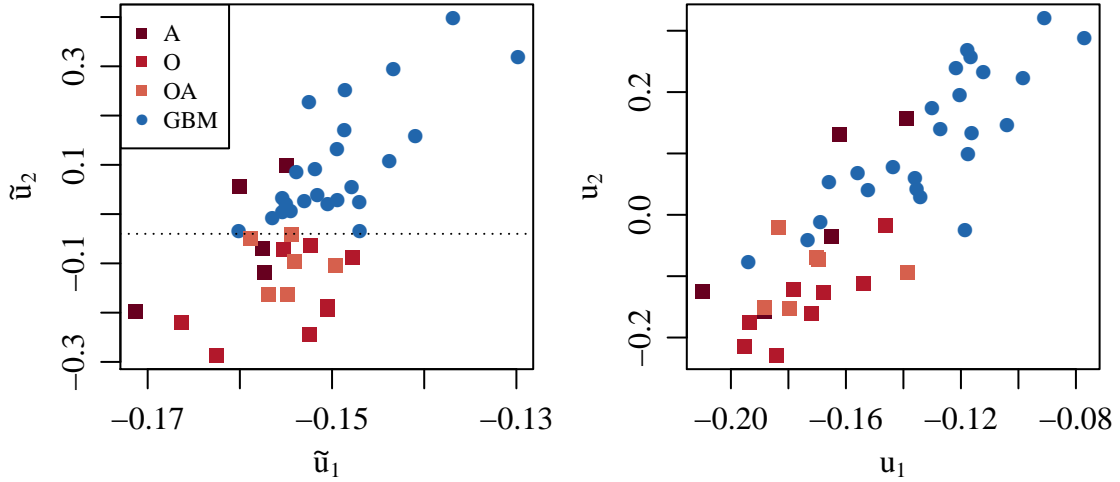


Figure 4: Plots of the first two principal components of  $\hat{\mathbf{M}}$  and  $\mathbf{Y}$ . Blue circles represent glioblastomas (GBM), while red squares represent astrocytomas (A), oligodendrogliomas (O) and mixed oligoastrocytomas (AO) which comprise the oligodendroglial tumors. The dashed line in the first panel represents a value of the second principal component of  $\hat{\mathbf{M}}$  that separates all 24 of the glioblastomas and two astrocytomas from the rest of the oligodendroglial tumors.

We examine the top ten strongest correlations between the second principal component,  $\tilde{\mathbf{u}}_2$ , and the columns of  $\hat{\mathbf{M}}$  to identify some of the genes that are differentially expressed in glioblastomas relative to oligodendroglial tumors. Table 1 shows the top ten genes most strongly correlated with  $\tilde{\mathbf{u}}_2$ , the value of the correlation, and the rank of the same gene with respect to the correlation between  $\mathbf{Y}$  and the second principal component of the raw data,  $\mathbf{u}_2$ .

	Corr. with $\tilde{\mathbf{u}}_2$	Rank of Corr. with $\mathbf{u}_2$
DLL3	-0.86	62
CKMT1A	-0.84	18
FBXL16	-0.83	1
AA281932	0.82	16
STXBP1	-0.81	11
PDXP	-0.80	2
X37864	-0.80	9
TIMP1	0.79	36
BGN	0.79	40
ABCC3	0.79	70

Table 1: Top ten strongest correlations between columns of  $\hat{\mathbf{M}}$  and  $\tilde{\mathbf{u}}_2$ .

Notably, DLL3, STXBP1 and TIMP1 are known to be differentially expressed in glioblastomas relative to oligodendroglial tumors (Bleeker et al., 2012; Dmitrenko et al., 2007; Groft et al., 2001). None of these genes are among the top ten most strongly correlated with  $\mathbf{u}_2$ , the principal component that best separates glioblastomas from oligodendroglial tumors computed from the raw data. Additionally, the gene BNIP2 has a much higher ranked association with the second principal component in the LANOVA denoised data than the raw data. BNIP2 was identified as associated with glioblastoma in Bredel et al. (2005) when this data originally appeared. This suggests that denoising microarray data using the LANOVA method may improve identification of differentially expressed genes.

**fMRI Data:** In our first example of LANOVA penalization for tensor data, we consider the problem of exploring how fMRI activations in response different tasks vary spatially in the brain. We use data which measures fMRI activations during 36 tasks for a single subject. This data originally appeared in Mitchell et al. (2004) and was also used to demonstrate a method for matrix decomposition in Allen et al. (2014). During each task, fMRI activations were measured at 55 time points and 4,698 spatial locations. Accordingly, the data can be represented as a  $36 \times 55 \times 4,698$  three-way tensor. The test given by Proposition 3.1 supports the use of a nonadditive estimate of  $\mathbf{M}$ ; we obtain a test statistic of 298.87 and reject the null hypothesis of normally distributed elementwise variability at level  $\alpha = 0.05$ . Having

found support for the use of a nonadditive estimate of  $\mathbf{M}$ , we also penalize lower-order mean parameters as  $\mathbf{Y}$  is high dimensional and sparsity of lower-order mean parameters could result in substantial dimension reduction and improved interpretability. The LANOVA penalized estimate has 1,751,179 nonzero parameters, a small fraction of the 9,302,040 parameters needed to represent the raw data,  $\mathbf{Y}$  (18.83%). Recalling that we are primarily interested in the relationship between task and spatial location, we examine the entries of  $\hat{\mathbf{F}}$  and  $\hat{\mathbf{C}}$ . Figure 5 shows the percent of nonzero entries of each at each spatial location. At each spatial location, the proportion of nonzero entries of  $\hat{\mathbf{F}}$  is much larger than the proportion of nonzero entries of  $\hat{\mathbf{C}}$ . This suggests that much of the spatial variation of activations by task is attributable to an overall level change in activation over the duration of the task, as opposed to time-specific changes in activation. For both  $\hat{\mathbf{F}}$  and  $\hat{\mathbf{C}}$ , we see spatial patterns in the percent of nonzero entries that give us some insight into the regions that are most responsive to the tasks performed in this data. By examining the percent of nonzero entries of  $\hat{\mathbf{F}}$  by spatial location, we can get a sense of which spatial locations correspond to level changes in fMRI activity response by task. There is evidence for an overall level change in response to at least some tasks for all voxels; the minimum percent of nonzero entries of  $\hat{\mathbf{F}}$  per spatial location is 33%. However, voxels in the the parietal region, the calcarine fissure and the right- and left-dorsolateral prefrontal cortex have particularly high proportions of nonzero entries of  $\hat{\mathbf{F}}$ , suggesting that overall activation in these regions is particularly responsive to tasks. By examining the percent of nonzero entries of  $\hat{\mathbf{C}}$  by spatial location, we can get a sense of which spatial locations correspond to time-specific differential activity by task. We see that nonzero entries of  $\hat{\mathbf{C}}$  are concentrated among voxels in the upper supplementary motor area, the calcarine fissure and the left- and right-temporal lobes. In this way, we can use LANOVA penalized estimates to inform future modeling decisions. Specifically, we can use LANOVA penalized estimates to identify subsets of relevant voxels and/or regions. Additionally, we can use LANOVA penalized estimates to assess how complex a model is needed to model spatial variation in activation responses to different tasks, e.g. to assess

when a relatively simple model with overall level changes in activation responses to different tasks at different locations is sufficient.

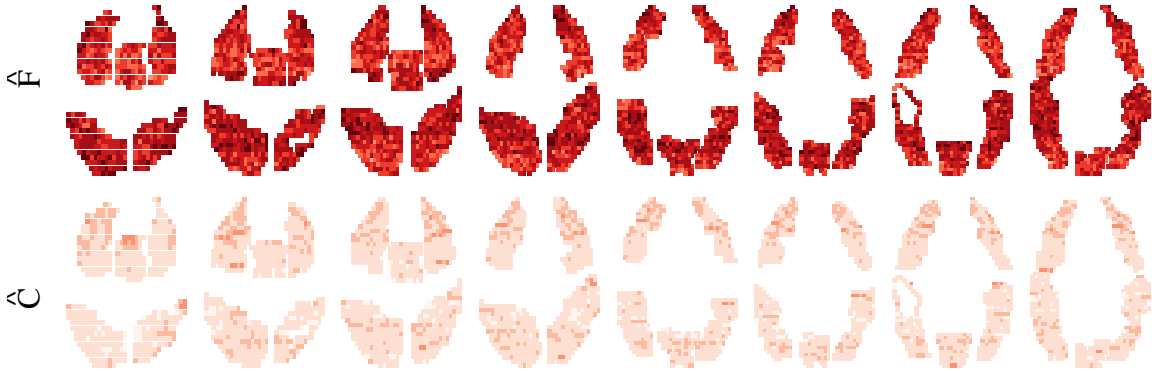


Figure 5: Percent of nonzero entries of  $\hat{\mathbf{F}}$  and  $\hat{\mathbf{C}}$  by spatial location. Darker colors indicate higher percentages.

**Fusarium Data:** In our last example, we consider the problem of checking for nonzero three way interactions in the experimental data without replicates. We examine experimental data measuring infection of wheat varieties, which has been used to demonstrate the use of multiplicative models for estimating elements of  $\mathbf{C}$  in the absence of replicate measurements (van Eeuwijk and Kroonenberg, 1998). The  $20 \times 7 \times 4$  three-way tensor contains severity of disease incidence ratings for 20 varieties of wheat infected with 7 strains of Fusarium head blight over 4 years, from 1990-1993. As in van Eeuwijk and Kroonenberg (1998), we transform the severity ratings to the logit scale before estimating LANOVA parameters. The test given by Proposition 3.1 supports the use of a nonadditive estimate of  $\mathbf{M}$ ; we obtain a test statistic of 3.99 and reject the null hypothesis of normally distributed elementwise variability at level  $\alpha = 0.05$ . When we estimate the unknown mean parameters, we obtain 87 nonzero entries of  $\hat{\mathbf{C}}$  (16%). Figure 6 shows nonzero elements of  $\hat{\mathbf{C}}$ , with dashed lines separating groups of wheat and blight by country of origin: Hungary (H), Germany (G), France (F) or the Netherlands (N). We can interpret elements of  $\hat{\mathbf{C}}$  as evidence of variety-by-year-by-strain interactions that cannot be expressed as additive in variety-by-year, year-by-strain and variety-by-strain effects. Like van Eeuwijk and Kroonenberg (1998), who identify



“real” three-way interactions using a single multiplicative component for  $\mathbf{C}$ , with  $c_{ijk} = \alpha_i \gamma_j \delta_k$ , we observe large three-way interactions in 1992, during which there was a disturbance in the storage of blight strains. Specifically, we observe interactions involving Dutch variety 2, the only variety with no infections at all in 1992, and interactions between Hungarian varieties 21 and 23 and foreign blight strains, which despite the storage disturbance were still able to cause infection in these two Hungarian varieties alone. We also observe evidence for “real” three way interactions for varieties exposed to Hungarian strain 12 in 1991 and French varieties exposed to Hungarian strains in 1993 as well as other interactions. We do not have enough information about the data to assess whether or not these estimated interactions are related to features of the study or known patterns in the behavior of certain varieties and strains, however they suggest further investigation of these varieties and strains may be warranted.

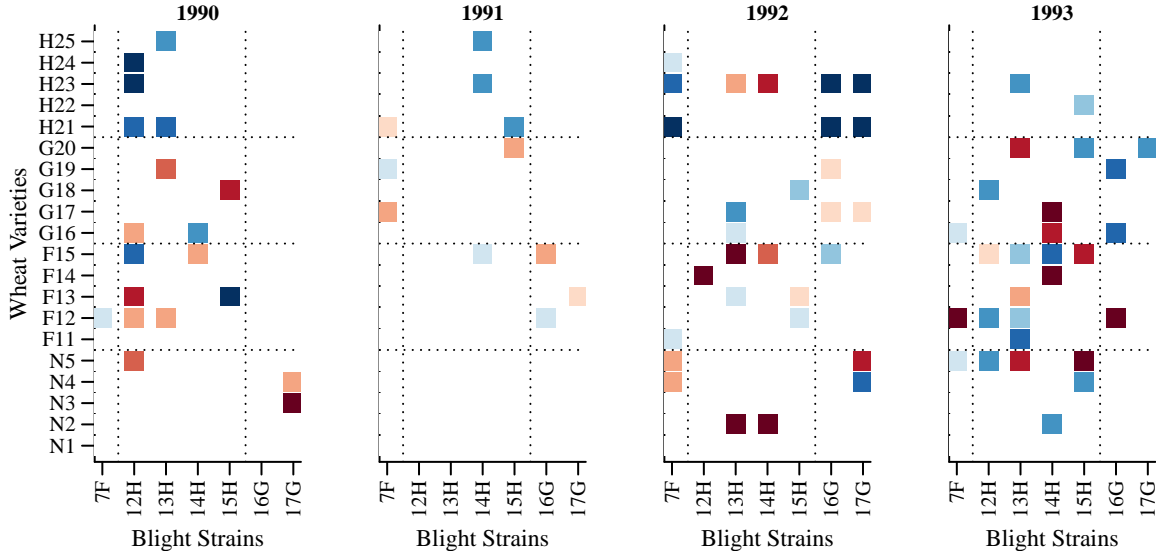


Figure 6: Entries of  $\hat{\mathbf{C}}$ . Red (blue) points indicate positive (negative) nonzero entries of  $\hat{\mathbf{C}}$  and darker colors correspond to larger magnitudes.

## 6 Discussion

Overall, this paper demonstrates that we can use the common Lasso penalty and the corresponding prior distribution to address the problem of estimating elementwise effects of scientific interest in the absence of replicates. In the matrix case, we can also think of LANOVA penalization as allowing us to assess evidence for nonadditivity. We show that our empirical Bayes estimators for the penalty parameters are consistent and we also show that the corresponding mean parameter estimates tend to be relatively robust to misspecification of the distribution of elements of  $\mathbf{C}$ . We emphasize that the empirical Bayes estimators are very simple and quick to compute for arbitrarily large matrices and tensors. Not only are they advantageous in the case where we penalize elements of  $\mathbf{C}$  alone, they provide a principled and fast approach to computing a large number of tuning parameters when we penalize lower-order mean parameters as well. Additionally, we note that the block coordinate descent algorithms we provide are made fast and efficient by exploiting the structure of the design matrix. Finally, we show that LANOVA penalized estimates can be used to improve classification by using LANOVA penalization to denoise data as a preprocessing step, to perform exploratory analysis of spatial variation in activation response to tasks in high dimensional fMRI data, and to assess evidence for the presence of “real” interaction effects at the elementwise level in experimental data collected using a factorial design.

The methods presented in this paper could be extended in several ways to address more general settings. Although we use our estimators of  $\lambda_c$  and  $\sigma_z^2$  in posterior mode estimation, the same estimators could also be used in posterior mean estimation. Focusing on posterior mean estimation of  $\mathbf{M}$  via a sampling based approach would be one way of addressing uncertainty of our estimate of  $\mathbf{M}$ , which we have largely ignored in this paper. Our empirical Bayes tuning parameter estimates could be used to set parameters of prior distributions for  $\lambda_c$  and/or  $\sigma_z^2$  in a more fully Bayesian approach. Additionally, LANOVA penalization for matrices is a specific case of the more general bilinear regression model, where we assume  $\mathbf{Y} = \mathbf{A}\mathbf{W} + \mathbf{B}\mathbf{X} + \mathbf{C} + \mathbf{Z}$ , given known  $\mathbf{W}$  and  $\mathbf{X}$ . The same logic used in this paper

to define estimators of  $\lambda_c$  and  $\sigma_z^2$  could be extended to this context and a more general multilinear context for tensor variate  $\mathbf{Y}$ .

Lastly, our own intuition as well as the results of Section 3 suggests that the Laplace distributional assumptions used in this paper are likely to be violated in many settings. Whereas the strength of the Laplace distributional assumption for elements of  $\mathbf{C}$  can be justified by the need to make *some* assumptions in the absence of replicate measurements, it cannot be similarly justified for lower-order mean parameters unless the specific application suggests it is appropriate. This leads us to one last extension of the material presented in this paper: improved penalty choice in the presence of replicate measurements. In future work, we will consider scenarios where we have enough information to estimate the variances of unknown mean parameters from second order information alone and use fourth order information to test the appropriateness of a penalty and/or choose a better one.

## References

- Allen, G. I., L. Grose, and J. Taylor (2014). A Generalized Least-Square Matrix Decomposition. *Journal of the American Statistical Association* 109(505), 145–159.
- Bleeker, F. E., R. J. Molenaar, and S. Leenstra (2012). Recent advances in the molecular understanding of glioblastoma. *Journal of Neuro-Oncology* 108(1), 11–27.
- Bredel, M., C. Bredel, D. Juric, G. R. Harsh, H. Vogel, L. D. Recht, and B. I. Sikic (2005). Functional Network Analysis Reveals Extended Gliomagenesis Pathway Maps and Three Novel MYC-Interacting Genes in Human Gliomas. *Cancer Research* 65(19), 8679–8689.
- Candes, E. J., C. A. Sing-Ling, and J. D. Trzasko (2013). Unbiased Risk Estimates for Singular Value Thresholding and Spectral Estimators. *IEEE Transactions on Signal Processing* 61(19), 4643–4657.
- de Tayrac, M., S. Lê, M. Aubry, J. Mosser, and F. Husson (2009). Simultaneous analysis of distinct Omics data sets with integration of biological knowledge: Multiple Factor Analysis approach. *BMC Genomics* 10(1), 32.
- Díaz-Francés, E. and J. A. Montoya (2008). Correction to On the linear combination of normal and Laplace random variables, by Nadarajah, S., Computational Statistics, 2006, 21, 6371. *Computational Statistics* 23(4), 661–666.
- Dmitrenko, V. V., O. I. Bojko, K. O. Shostak, N. Y. Vitak, T. V. Bukreeva, V. D. Rozumenko, T. A. Malysheva, M. I. Shamayev, Y. P. Zozulya, and V. M. Kavsan (2007).

- Characterization of genes, down-regulated in human glioma, potential tumor suppressor genes. *Biopolymers and Cell* 23(4), 347–362.
- Fazel, M. (2002). *Matrix Rank Minimization with Applications*. Phd thesis, Stanford University.
- Figueiredo, M. A. T. (2003). Adaptive Sparseness for Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(9), 1150–1159.
- Gerard, D. and P. Hoff (2015). Adaptive Higher-order Spectral Estimators.
- Gollob, H. F. (1968). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* 33(1), 73–115.
- Goodman, L. A. and S. J. Haberman (1990). The Analysis of Nonadditivity in Two-Way Analysis of Variance. *Journal of the American Statistical Association* 85(409), 139–145.
- Groft, L. L., H. Muzik, N. B. Rewcastle, R. N. Johnston, V. Knäuper, M. A. Lafleur, P. A. Forsyth, and D. R. Edwards (2001). Differential expression and localization of TIMP-1 and TIMP-4 in human gliomas. *British Journal of Cancer* 85(1), 55–63.
- Hoff, P. D. (2007). Model averaging and dimension selection for the singular value decomposition. *Journal of the American Statistical Association* 102(478), 674–685.
- Johnson, D. E. and F. A. Graybill (1972). An Analysis of a Two-Way Model with Interaction and No Replication. *Journal of the American Statistical Association* 67(340), 862–868.
- Josse, J. and S. Sardy (2016). Adaptive shrinkage of singular values. *Statistics and Computing* 26(3), 715–724.
- Josse, J., S. Sardy, and S. Wager (2016). denoiseR: A Package for Low Rank Matrix Estimation.
- Konig, R. H., H. Neudecker, and T. Wansbeek (1992). Unbiased Estimation of Fourth-Order Matrix Moments. *Linear Algebra and Its Applications* 160, 163–174.
- Kotz, S., T. J. Kuzobowski, and K. Podgórski (2001). *The Laplace Distribution and Generalizations*. Boston: Birkhäuser.
- Mandel, J. (1971). A New Analysis of Variance Model for Non-Additive Data. *Technometrics* 13(1), 1–18.
- Mitchell, T. M., R. Hutchinson, R. S. Niculescu, and X. Wang (2004). Learning to Decode Cognitive States from Brain Images. *Machine Learning* 57(January), 145–175.
- Nadarajah, S. (2006). On the linear combination of normal and Laplace random variables. *Computational Statistics* 21(1), 63–71.
- Park, T. and G. Casella (2008). The Bayesian Lasso. *Journal of the American Statistical Association* 103(482), 681–686.

- Tibshirani, R. J. (2013). The lasso problem and uniqueness. *Electronic Journal of Statistics* 7(1), 1456–1490.
- van Eeuwijk, F. A. and P. M. Kroonenberg (1998). Multiplicative Models for Interaction in Three-Way ANOVA, with Applications to Plant Breeding. *Biometrics* 54(4), 1315–1333.

## Appendix

**Uniqueness of  $\hat{\mathbf{M}}$ ,  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{C}}$ :** As this is a standard lasso regression problem, uniqueness of  $\hat{\mathbf{M}}$  is well known Tibshirani (2013). Suppose we have two unique solutions for  $\mathbf{A}$ :  $\hat{\mathbf{A}}$  and  $\tilde{\mathbf{A}}$ , where  $\hat{\mathbf{A}} \neq \tilde{\mathbf{A}}$ . Then it must be the case that  $\tilde{\mathbf{A}} = \hat{\mathbf{A}} + \boldsymbol{\alpha}\mathbf{1}'_p + \mathbf{1}_n\boldsymbol{\beta}'$ , where  $\boldsymbol{\alpha} \neq \mathbf{0}$  and/or  $\boldsymbol{\beta} \neq \mathbf{0}$ , i.e.  $\boldsymbol{\alpha}\mathbf{1}'_p + \mathbf{1}_n\boldsymbol{\beta}' \neq \mathbf{0}$ . Let  $\hat{\mathbf{C}}$  and  $\tilde{\mathbf{C}}$  be the corresponding solutions for  $\mathbf{C}$ . By uniqueness of  $\hat{\mathbf{M}}$ , we have:

$$\hat{\mathbf{A}} + \hat{\mathbf{C}} = \tilde{\mathbf{A}} + \tilde{\mathbf{C}} = \hat{\mathbf{M}}.$$

It follows that  $\hat{\mathbf{C}} = \hat{\mathbf{M}} - \hat{\mathbf{A}}$  and  $\tilde{\mathbf{C}} = \hat{\mathbf{M}} - \tilde{\mathbf{A}}$ . Substituting our expression for  $\tilde{\mathbf{A}}$  in terms of  $\hat{\mathbf{A}}$  into  $\tilde{\mathbf{C}}$  yields  $\tilde{\mathbf{C}} = \hat{\mathbf{M}} - \hat{\mathbf{A}} - \boldsymbol{\alpha}\mathbf{1}'_p - \mathbf{1}_n\boldsymbol{\beta}'$ . Then:

$$\begin{aligned} \left\| \text{vec}(\tilde{\mathbf{C}}) \right\|_1 &= \left\| \text{vec}(\hat{\mathbf{M}} - \hat{\mathbf{A}} - \boldsymbol{\alpha}\mathbf{1}'_p - \mathbf{1}_n\boldsymbol{\beta}') \right\|_1 \\ &\leq \left\| \text{vec}(\hat{\mathbf{M}} - \hat{\mathbf{A}}) \right\|_1 + \left\| \text{vec}(\boldsymbol{\alpha}\mathbf{1}'_p + \mathbf{1}_n\boldsymbol{\beta}') \right\|_1 \\ &= \left\| \text{vec}(\hat{\mathbf{C}}) \right\|_1 + \left\| \text{vec}(\boldsymbol{\alpha}\mathbf{1}'_p + \mathbf{1}_n\boldsymbol{\beta}') \right\|_1. \end{aligned}$$

It follows that  $\left\| \text{vec}(\tilde{\mathbf{C}}) \right\|_1 < \left\| \text{vec}(\hat{\mathbf{C}}) \right\|_1$  and

$$\frac{1}{2\sigma_z^2} \left\| \text{vec}(\mathbf{Y} - \hat{\mathbf{M}}) \right\|_2^2 + \lambda_c \left\| \text{vec}(\tilde{\mathbf{C}}) \right\|_1 < \frac{1}{2\sigma_z^2} \left\| \text{vec}(\mathbf{Y} - \hat{\mathbf{M}}) \right\|_2^2 + \lambda_c \left\| \text{vec}(\hat{\mathbf{C}}) \right\|_1.$$

This contracts our claim that the  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{C}}$  solve the penalized regression problem. It follows that  $\hat{\mathbf{A}}$  is unique and as  $\hat{\mathbf{C}}$  is uniquely determined by  $\hat{\mathbf{M}}$  and  $\hat{\mathbf{A}}$ , it follows that  $\hat{\mathbf{C}}$  is unique as well.

**Equivalence of ANOVA Decomposition of  $\mathbf{Y}$  and  $\hat{\mathbf{M}}$ :** We can rewrite the penalized problem as follows:

$$\min_{\mu, \mathbf{a}, \mathbf{b}, \mathbf{C}} \frac{1}{2\sigma_z^2} \left\| \text{vec}(\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) \right\|_2^2 + \lambda_c \left\| \text{vec}(\mathbf{C}) \right\|_1,$$

where  $\boldsymbol{\theta}' = [\mu \quad \mathbf{a}' \quad \mathbf{b}' \quad \text{vec}(\mathbf{C})']$  and

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_p \otimes \mathbf{1}_n & \mathbf{1}_p \otimes \mathbf{I}_n & \mathbf{I}_p \otimes \mathbf{1}_n & \mathbf{I}_p \otimes \mathbf{I}_n \end{bmatrix}. \quad (8)$$

A solution,  $\hat{\boldsymbol{\theta}} = [\hat{\mu} \quad \hat{\mathbf{a}} \quad \hat{\mathbf{b}} \quad \text{vec}(\hat{\mathbf{C}})]'$  satisfies the KKT conditions:

$$\mathbf{X}' \left( \text{vec}(\mathbf{Y} - \hat{\mathbf{M}}) \right) = \begin{bmatrix} 0 \cdot \mathbf{1}_{1+n+p} \\ \lambda_c \text{vec}(\mathbf{\Gamma}) \end{bmatrix},$$

where  $\gamma_{ij} \in \begin{cases} \{\text{sign}(\hat{c}_{ij})\} & \text{if } \hat{c}_{ij} = 0 \\ [-1, 1] & \text{if } \hat{c}_{ij} \neq 0 \end{cases}$ .

Using Equation (8), the first  $1 + n + p$  equalities give:

$$\begin{aligned}\bar{y}_{..} &= \bar{\tilde{m}}_{..} \\ \bar{y}_{i.} &= \bar{\tilde{m}}_{i.}, \quad i = 1, \dots, n \\ \bar{y}_{.j} &= \bar{\tilde{m}}_{.j}, \quad j = 1, \dots, p,\end{aligned}$$

where  $\bar{y}_{..} = \frac{1}{np} \sum_{i=1}^n \sum_{j=1}^p y_{ij}$ ,  $\bar{y}_{i.} = \frac{1}{p} \sum_{j=1}^p y_{ij}$  and  $\bar{y}_{.j} = \frac{1}{n} \sum_{i=1}^n y_{ij}$  and  $\bar{\tilde{m}}_{..}$ ,  $\bar{\tilde{m}}_{i.}$  and  $\bar{\tilde{m}}_{.j}$  are defined accordingly. This means that the two-way ANOVA decompositions fits based on  $\mathbf{Y}$  and  $\hat{\mathbf{M}}$  will be the same.

**Proof of Proposition 2.1:** To prove this proposition, we need to compute  $\mathbb{E}[\bar{r}^{(4)}]$  and  $\mathbb{E}[(\bar{r}^{(2)})^2]$ . We use the following result from Konig et al. (1992):

$$\mathbb{E}[(\mathbf{x}'\mathbf{S}\mathbf{x})^2] = \alpha \mathbb{E}[x_i^4] + (\theta + 2\beta) \mathbb{E}[x_i^2 x_j^2], \quad (9)$$

for a mean zero random vector  $\mathbf{x}$  of independent elements and symmetric fixed matrix  $\mathbf{S}$ , where  $\alpha = \sum_i s_{ii}^2$ ,  $\theta = \sum_{i \neq j} s_{ii} s_{jj}$  and  $\beta = \sum_{i \neq j} s_{ij}^2$ .

First, we rewrite  $\mathbb{E}[(\bar{r}^{(2)})^2]$  to resemble Equation (9):

$$(\bar{r}^{(2)})^2 = \frac{1}{n^2 p^2} (\text{vec}(\mathbf{C} + \mathbf{Z})' (\mathbf{H}_p \otimes \mathbf{H}_n) (\text{vec}(\mathbf{C} + \mathbf{Z})))^2.$$

We compute  $\alpha$ ,  $\theta$  and  $\beta$  below:

$$\begin{aligned}\alpha &= (np) \left( \frac{(n-1)(p-1)}{np} \right)^2 \\ &= \frac{(n-1)^2 (p-1)^2}{np} \\ \theta &= np(np-1) \left( \frac{(n-1)(p-1)}{np} \right)^2 \\ &= \frac{(np-1)(n-1)^2 (p-1)^2}{np} \\ \beta &= np(p-1) \left( \frac{n-1}{np} \right)^2 + n(n-1)p \left( \frac{p-1}{np} \right)^2 + n(n-1)p(p-1) \left( \frac{1}{np} \right)^2 \\ &= \frac{(n-1)^2 (p-1) + (n-1)(p-1)^2 + (n-1)(p-1)}{np}\end{aligned}$$

Recalling  $\kappa$  is defined such that  $\mathbb{E}[c_{ij}^4] = (\kappa + 3) \sigma_c^4$  and plugging these expressions for

$\alpha$ ,  $\beta$  and  $\theta$  into Equation (9) yields:

$$\begin{aligned}
\mathbb{E} \left[ (\bar{r}^{(2)})^2 \right] &= \frac{(n-1)^2 (p-1)^2}{n^3 p^3} \left( \kappa \sigma_c^4 + 3 (\sigma_c^2 + \sigma_z^2)^2 \right) + \\
&\quad \frac{(np-1) (n-1)^2 (p-1)^2}{n^3 p^3} (\sigma_c^2 + \sigma_z^2)^2 + \\
&\quad 2 \left( \frac{(n-1)^2 (p-1) + (n-1) (p-1)^2 + (n-1) (p-1)}{n^3 p^3} \right) (\sigma_c^2 + \sigma_z^2)^2 \\
&= \kappa \left( \frac{(n-1)^2 (p-1)^2}{n^3 p^3} \right) \sigma_c^4 + \left( \frac{(n-1) (p-1)}{np} \right)^2 (\sigma_c^2 + \sigma_z^2)^2 + \\
&\quad 2 \left( \frac{(n-1) (p-1)}{n^2 p^2} \right) (\sigma_c^2 + \sigma_z^2)^2.
\end{aligned} \tag{10}$$

Now we compute  $\bar{r}^{(4)}$ . Letting  $\mathbf{E}_{kk}$  be an  $np \times np$  matrix with the  $k$ -th diagonal element equal to 1 and all other elements equal to 0. We can rewrite  $\bar{r}^{(4)}$  so each term in resembles Equation (9):

$$\begin{aligned}
\bar{r}^{(4)} &= \frac{1}{np} \sum_{k=1}^{np} \left( \text{vec}(\mathbf{R})' \mathbf{E}_{kk} \text{vec}(\mathbf{R}) \right)^2 \\
&= \frac{1}{np} \sum_{k=1}^{np} \left( \text{vec}(\mathbf{C} + \mathbf{Z})' (\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n) \text{vec}(\mathbf{C} + \mathbf{Z}) \right)^2.
\end{aligned}$$

The matrices  $(\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n)$  take the place of  $\mathbf{S}$  in Equation (9).

From Equation (9), we have to compute the sum of squared diagonal elements, denoted by  $\alpha_k$ , the sum of pairwise products of distinct diagonal elements, denoted by  $\theta_k$ , and the sum of squared off-diagonal elements, denoted by  $\beta_k$ , for each of these matrices. Before going further, we show that  $\theta_k = \beta_k$ . Letting  $\mathbf{e}_k$  is an  $np \times 1$  vector with the  $k$ -th element equal



to 1 and all other elements equal to 0,

$$\begin{aligned}
\alpha_k + \beta_k &= \text{tr}((\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n) (\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n)) \\
&= \text{tr}((\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{e}_k \mathbf{e}_k' (\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{e}_k \mathbf{e}_k') \\
&= (\mathbf{e}_k' (\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{e}_k)^2 \\
&= \left( \frac{(n-1)(p-1)}{np} \right)^2, \\
\alpha_k + \theta_k &= \text{tr}((\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n) \otimes (\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n)) \\
&= \text{tr}((\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n))^2 \\
&= \text{tr}((\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{e}_k \mathbf{e}_k')^2 \\
&= (\mathbf{e}_k' (\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{e}_k)^2 \\
&= \left( \frac{(n-1)(p-1)}{np} \right)^2.
\end{aligned}$$

It follows that  $\beta_k = \theta_k$ .

This means we *only* need to compute the sum of squared diagonal elements and the sum of pairwise products of distinct diagonal elements of  $(\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n)$ .

Each matrix,  $(\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n)$ , is constructed by multiplying the  $k$ -th column and the  $k$ -th row of  $(\mathbf{H}_p \otimes \mathbf{H}_n)$ . By symmetry of  $(\mathbf{H}_p \otimes \mathbf{H}_n)$ , this is the same as multiplying the  $k$ -th column, denoted by  $(\mathbf{H}_p \otimes \mathbf{H}_n)_k$ , by its transpose,

$$(\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n) = (\mathbf{H}_p \otimes \mathbf{H}_n)_k (\mathbf{H}_p \otimes \mathbf{H}_n)_k'.$$

Regardless of  $k$ , each column  $(\mathbf{H}_p \otimes \mathbf{H}_n)_k$  has the following elements:

- One element equal to  $\frac{(n-1)(p-1)}{np}$ ;
- $(p-1)$  elements equal to  $-\frac{(n-1)}{np}$ ;
- $(n-1)$  elements equal to  $-\frac{(p-1)}{np}$ ;
- $(n-1)(p-1)$  elements equal to  $\frac{1}{np}$ .

The diagonal elements of  $(\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n)$  will be given by the squared elements of  $(\mathbf{H}_p \otimes \mathbf{H}_n)_k$ . Note that the only feature of the diagonal elements of  $(\mathbf{H}_p \otimes \mathbf{H}_n) \mathbf{E}_{kk} (\mathbf{H}_p \otimes \mathbf{H}_n)$  that depends on  $k$  is the order of these elements. Accordingly, the sum of squared diagonal elements,  $\alpha_k$ , and the sum of pairwise products of distinct diagonal elements,  $\theta_k$ , will *not*

depend on  $k$ . Dropping the subscript, we compute  $\alpha$  and  $\theta$ :

$$\begin{aligned}
\alpha &= \left( \frac{(n-1)(p-1)}{np} \right)^4 + (p-1) \left( \frac{n-1}{np} \right)^4 + (n-1) \left( \frac{p-1}{np} \right)^4 + (n-1)(p-1) \left( \frac{1}{np} \right)^4 \\
&= \frac{(n-1)(p-1)}{n^4 p^4} ((n-1)^3 (p-1)^3 + (n-1)^3 + (p-1)^3 + 1) \\
&= \frac{(n-1)(n^2 - 3n + 3)(p-1)(p^2 - 3p + 3)}{n^4 p^4} \\
\theta &= (p-1) \left( \frac{(n-1)^4 (p-1)^2}{n^4 p^4} \right) + (n-1) \left( \frac{(n-1)^2 (p-1)^4}{n^4 p^4} \right) + (n-1)(p-1) \left( \frac{(n-1)^2 (p-1)^2}{n^4 p^4} \right) + \\
&\quad (p-1)(p-2) \left( \frac{(n-1)^4}{n^4 p^4} \right) + (n-1)(p-1) \left( \frac{(n-1)^2 (p-1)^2}{n^4 p^4} \right) + (n-1)(p-1)^2 \left( \frac{(n-1)^2}{n^4 p^4} \right) + \\
&\quad (n-1)(n-2) \left( \frac{(p-1)^4}{n^4 p^4} \right) + (n-1)^2 (p-1) \left( \frac{(p-1)^2}{n^4 p^4} \right) + \\
&\quad (n-1)(p-1)((n-1)(p-1)-1) \left( \frac{1}{n^4 p^4} \right) \\
&= \frac{1}{n^4 p^4} ((p-1)^3 (n-1)^4 + (p-1)^4 (n-1)^3 + (p-1)^3 (n-1)^3 + \\
&\quad + (p-1)^3 (n-1)^4 + (p-1)(p-2)(n-1)^4 + (p-1)^3 (n-1)^3 + (p-1)^2 (n-1)^3 + \\
&\quad + (p-1)^4 (n-1)^3 + (p-1)^3 (n-1)^3 + (n-1)(n-2)(p-1)^4 + (p-1)^3 (n-1)^2 + \\
&\quad + (n-1)^3 (p-1)^3 + (n-1)^3 (p-1)^2 + (n-1)^2 (p-1)^3 + (n-1)(p-1)((n-1)(p-1)-1)) \\
&= \frac{1}{n^4 p^4} (2(p-1)^3 (n-1)^4 + 2(p-1)^4 (n-1)^3 + 4(p-1)^3 (n-1)^3 + \\
&\quad + (p-1)(p-2)(n-1)^4 + 2(p-1)^2 (n-1)^3 + (n-1)(n-2)(p-1)^4 + 2(p-1)^3 (n-1)^2 + \\
&\quad + (n-1)(p-1)((n-1)(p-1)-1)) \\
&= \frac{(n-1)(p-1)(2np^2 + 2n^2p - 3p^2 - 3n^2 - 8np + 9p + 9n - 9)}{n^3 p^3} \\
&= \left( \frac{(n-1)(p-1)}{np} \right)^2 - \frac{(n-1)(n^2 - 3n + 3)(p-1)(p^2 - 3p + 3)}{n^3 p^3}.
\end{aligned}$$

Plugging these expressions into Equation (9) yields:

$$\begin{aligned}
\mathbb{E} [\bar{r}^{(4)}] &= \kappa \left( \frac{(n-1)(n^2 - 3n + 3)(p-1)(p^2 - 3p + 3)}{n^3 p^3} \right) \sigma_c^4 + \\
&\quad 3 \left( \frac{(n-1)(p-1)}{np} \right)^2 (\sigma_c^2 + \sigma_z^2)^2.
\end{aligned} \tag{11}$$

Plugging in the excess kurtosis,  $\kappa = 3$ , of Laplace distributed  $c_{ij}$ , Proposition 2.1 follows straightforwardly from the definition of  $\hat{\sigma}_c^4$  and Equations (10) and (11).

**Proof of Proposition 2.2:** We start by showing consistency of our empirical Bayes estimators as  $p \rightarrow \infty$  with  $n$  fixed. The proof of consistency of our empirical Bayes estimators as  $n \rightarrow \infty$  with  $p$  fixed is analogous.

Recalling that our empirical Bayes estimators are functions of  $\bar{r}^{(2)}$  and  $\bar{r}^{(4)}$ , we start by proving that  $\bar{r}^{(2)} \xrightarrow{p} \left(\frac{n-1}{n}\right) (\sigma_c^2 + \sigma_z^2)$  as  $p \rightarrow \infty$  for fixed  $n$ . First, we compute  $\mathbb{E} [\bar{r}^{(2)}]$ .

$$\begin{aligned}
\mathbb{E} [\bar{r}^{(2)}] &= \frac{1}{np} \mathbb{E} [\text{vec}(\mathbf{R})' \text{vec}(\mathbf{R})] \\
&= \frac{1}{np} \mathbb{E} [\text{vec}(\mathbf{C} + \mathbf{Z})' (\mathbf{H}_p \otimes \mathbf{H}_n) \text{vec}(\mathbf{C} + \mathbf{Z})] \\
&= \frac{1}{np} \mathbb{E} [\text{vec}(\mathbf{C} + \mathbf{Z})' (\mathbf{H}_p \otimes \mathbf{H}_n) \text{vec}(\mathbf{C} + \mathbf{Z})] \\
&= \frac{1}{np} \text{tr} ((\mathbf{H}_p \otimes \mathbf{H}_n) \mathbb{E} [\text{vec}(\mathbf{C} + \mathbf{Z}) \text{vec}(\mathbf{C} + \mathbf{Z})']) \\
&= \frac{1}{np} \text{tr} (\mathbf{H}_p \otimes \mathbf{H}_n) (\sigma_c^2 + \sigma_z^2) \\
&= \frac{(n-1)(p-1)}{np} (\sigma_c^2 + \sigma_z^2).
\end{aligned} \tag{12}$$

For fixed  $n$ , as  $p \rightarrow \infty$ ,  $\mathbb{E} [\bar{r}^{(2)}] = O(1) \left(\frac{n-1}{n}\right) (\sigma_c^2 + \sigma_z^2)$ .

We can compute  $\mathbb{V} [\bar{r}^{(2)}]$  explicitly using Equation (10).

$$\mathbb{V} [\bar{r}^{(2)}] = \kappa \left( \frac{(n-1)^2 (p-1)^2}{n^3 p^3} \right) \sigma_c^4 + 2 \left( \frac{(n-1)(p-1)}{n^2 p^2} \right) (\sigma_c^2 + \sigma_z^2)^2. \tag{13}$$

For fixed  $n$ , as  $p \rightarrow \infty$ ,  $\mathbb{V} [\bar{r}^{(2)}] = O\left(\frac{1}{p}\right)$ . It follows that  $\bar{r}^{(2)} \xrightarrow{p} \left(\frac{n-1}{n}\right) (\sigma_c^2 + \sigma_z^2)$  as  $p \rightarrow \infty$  for fixed  $n$ .

Now we show  $\bar{r}^{(4)} \xrightarrow{p} \kappa \left( \frac{(n-1)(n^2-3n+3)}{n^3} \right) \sigma_c^4 + 3 \left( \frac{n-1}{n} \right)^2 (\sigma_c^2 + \sigma_z^2)^2$  as  $p \rightarrow \infty$  for fixed  $n$ .

From Equation (11), we have  $\mathbb{E} [\bar{r}^{(4)}] = \left( O(1) \kappa \frac{(n-1)(n^2-3n+3)}{n^3} \right) \sigma_c^4 + O(1) 3 \left( \frac{n-1}{n} \right)^2 (\sigma_c^2 + \sigma_z^2)^2$

as  $p \rightarrow \infty$  for fixed  $n$ . Now we evaluate the order of  $\mathbb{V} [\bar{r}^{(4)}]$ , starting by evaluating the order of  $\mathbb{E} \left[ \left( \bar{r}^{(4)} \right)^2 \right]$ . For convenience, we let  $x_{ij} = c_{ij} + z_{ij}$  and use the following way of writing elements of  $\mathbf{R}$ :  $r_{ij} = x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}$ , where  $\bar{x}_{i.} = \frac{1}{p} \sum_{j=1}^p x_{ij}$ ,  $\bar{x}_{.j} = \frac{1}{n} \sum_{i=1}^n x_{ij}$  and  $\bar{x}_{..} = \frac{1}{np} \sum_{i=1}^n \sum_{j=1}^p x_{ij}$ .

$$\begin{aligned}
\mathbb{E} \left[ (\bar{r}^{(4)})^2 \right] &= \mathbb{E} \left[ \left( \frac{1}{np} \sum_{i=1}^n \sum_{j=1}^p r_{ij}^2 \right)^2 \right] \\
&= \frac{1}{n^2 p^2} \mathbb{E} \left[ \sum_{i=1}^n \sum_{j=1}^p r_{ij}^8 + \sum_{i'=1}^n \sum_{j'=p, i'j' \neq ij} r_{ij}^4 r_{i'j'}^4 \right] \\
&= \frac{1}{n^2 p^2} \mathbb{E} \left[ \sum_{i=1}^n \sum_{j=1}^p \left( x_{ij} - \bar{x}_{\cdot j} + O_p \left( \frac{1}{\sqrt{p}} \right) \right)^8 \right] + \\
&\quad \frac{1}{n^2 p^2} \mathbb{E} \left[ \sum_{i=1}^n \sum_{j=1}^p \sum_{i' \neq i} \left( x_{ij} - \bar{x}_{\cdot j} + O_p \left( \frac{1}{\sqrt{p}} \right) \right)^4 \left( x_{i'j'} - \bar{x}_{\cdot j'} + O_p \left( \frac{1}{\sqrt{p}} \right) \right)^4 \right] + \\
&\quad \frac{1}{n^2 p^2} \mathbb{E} \left[ \sum_{i=1}^n \sum_{j=1}^p \sum_{i'=1}^n \sum_{j' \neq j} \left( x_{ij} - \bar{x}_{\cdot j} + O_p \left( \frac{1}{\sqrt{p}} \right) \right)^4 \left( x_{i'j'} - \bar{x}_{\cdot j'} + O_p \left( \frac{1}{\sqrt{p}} \right) \right)^4 \right] \\
&= \frac{1}{np} \mathbb{E} [x_{11}^8] + O \left( \frac{1}{p} \right) + \\
&\quad \frac{1}{n^2 p^2} \sum_{i=1}^n \sum_{j=1}^p \sum_{i'=1}^n \mathbb{E} [(x_{ij} - \bar{x}_{\cdot j})^4 (x_{i'j} - \bar{x}_{\cdot j})^4] + O \left( \frac{1}{\sqrt{p}} \right) + \\
&\quad \frac{1}{n^2 p^2} \sum_{i=1}^n \sum_{j=1}^p \sum_{i'=1}^n \sum_{j' \neq j} \mathbb{E} [(x_{ij} - \bar{x}_{\cdot j})^4] \mathbb{E} [(x_{i'j'} - \bar{x}_{\cdot j'})^4] + O \left( \frac{1}{p^2} \right) \\
&= \frac{1}{p} \underbrace{\mathbb{E} [(x_{11} - \bar{x}_{\cdot 1})^4 (x_{21} - \bar{x}_{\cdot 1})^4]}_* + \left( \frac{p-1}{p} \right) \mathbb{E} [(x_{11} - \bar{x}_{\cdot 1})^4] \mathbb{E} [(x_{12} - \bar{x}_{\cdot 2})^4] + O \left( \frac{1}{\sqrt{p}} \right) \\
&= \left( \frac{p-1}{p} \right) \mathbb{E} [(x_{11} - \bar{x}_{\cdot 1})^4]^2 + O \left( \frac{1}{\sqrt{p}} \right).
\end{aligned}$$

The term marked by  $*$  is constant for fixed  $n$ , it is a degree eight polynomial of  $n$  independent elements of  $\mathbf{X}$ , with coefficients that depend on  $n$  alone.

Now we need to compute the remaining unknown term,

$$\begin{aligned}
\mathbb{E}[(x_{11} - \bar{x}_{.1})^4] &= \mathbb{E}\left[\left(\frac{n-1}{n}x_{11} - \frac{1}{n}\sum_{i=2}^n x_{i1}\right)^4\right] \\
&= \left(\frac{n-1}{n}\right)^4 \mathbb{E}[x_{11}^4] + 6\left(\frac{n-1}{n}\right)^2 \mathbb{E}[x_{11}^2] \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=2}^n x_{i1}\right)^2\right] + \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=2}^n x_{i1}\right)^4\right] \\
&= \left(\frac{n-1}{n}\right)^4 \mathbb{E}[x_{11}^4] + \frac{6(n-1)^3}{n^4} \mathbb{E}[x_{11}^2]^2 + \frac{n-1}{n^4} \mathbb{E}[x_{11}^4] + \frac{3(n-1)(n-2)}{n^4} \mathbb{E}[x_{11}^2]^2 \\
&= \frac{(n-1)((n-1)^3 + 1)}{n^4} \mathbb{E}[x_{11}^4] + \frac{(n-1)(6(n-1)^2 + 3(n-2))}{n^4} \mathbb{E}[x_{11}^2]^2 \\
&= \frac{(n-1)(n^2 - 3n + 3)}{n^3} \mathbb{E}[x_{11}^4] + \frac{3(n-1)(2n-3)}{n^3} \mathbb{E}[x_{11}^2]^2 \\
&= \kappa \frac{(n-1)(n^2 - 3n + 3)}{n^3} \sigma_c^4 + 3\left(\frac{n-1}{n}\right)^2 (\sigma_c^2 + \sigma_z^2)^2.
\end{aligned}$$

Combining this with our expression for  $\mathbb{E}[(\bar{r}^{(4)})^2]$ , we have  $\mathbb{V}[(\bar{r}^{(4)})^2] = O\left(\frac{1}{\sqrt{p}}\right)$  as  $p \rightarrow \infty$  for fixed  $n$ . It follows that  $\bar{r}^{(4)} \xrightarrow{p} \kappa \left(\frac{(n-1)(n^2-3n+3)}{n^3}\right) \sigma_c^4 + 3\left(\frac{n-1}{n}\right)^2 (\sigma_c^2 + \sigma_z^2)^2$  as  $p \rightarrow \infty$  for fixed  $n$ .

Plugging in the excess kurtosis,  $\kappa = 3$ , of Laplace distributed  $c_{ij}$ , combining  $\bar{r}^{(4)} \xrightarrow{p} 3\left(\frac{(n-1)(n^2-3n+3)}{n^3}\right) \sigma_c^4 + 3\left(\frac{n-1}{n}\right)^2 (\sigma_c^2 + \sigma_z^2)^2$  as  $p \rightarrow \infty$  for fixed  $n$  with  $\bar{r}^{(2)} \xrightarrow{p} \left(\frac{n-1}{n}\right) (\sigma_c^2 + \sigma_z^2)$  as  $p \rightarrow \infty$  for fixed  $n$ , and applying the continuous mapping theorem yields  $\hat{\sigma}_c^4 \xrightarrow{p} \sigma_c^4$ ,  $\hat{\sigma}_c^2 \xrightarrow{p} \sigma_c^2$ ,  $\hat{\lambda}_c \xrightarrow{p} \lambda_c$  and  $\hat{\sigma}_z^2 \xrightarrow{p} \sigma_z^2$  as  $p \rightarrow \infty$  for fixed  $n$ .

Now we show consistency of our empirical Bayes estimators as  $n, p \rightarrow \infty$ . It follows directly from Equations (17) and (13) that  $\bar{r}^{(2)} \xrightarrow{p} \sigma_c^2 + \sigma_z^2$  as  $n, p \rightarrow \infty$ . Regarding  $\bar{r}^{(4)}$ , we again let  $x_{ij} = c_{ij} + z_{ij}$  and use the following decomposition:

$$r_{ij} = x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}_{\dots}$$

For convenience, we also let  $n = \pi_1 q$ ,  $p = \pi_2 q$ , where  $0 < \pi_1 < 1$  and  $0 < \pi_2 < 1$  are fixed and  $\pi_1 + \pi_2 = 1$ . Then the scenario  $n, p \rightarrow \infty$  is equivalent to  $q \rightarrow \infty$ . We have:

$$\begin{aligned}
\bar{r}^{(4)} &= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^4 \\
&= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \left( x_{ij} + O_p \left( \frac{1}{\sqrt{q}} \right) \right)^4 \\
&= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^4 + x_{ij}^3 O_p \left( \frac{1}{\sqrt{q}} \right) + x_{ij}^2 O_p \left( \frac{1}{q} \right) + x_{ij} O_p \left( \frac{1}{\sqrt{q}^3} \right) + O_p \left( \frac{1}{q^2} \right) \\
&= \underbrace{\bar{x}^{(4)} - \left( \kappa \sigma_c^4 + 3 (\sigma_c^2 + \sigma_z^2)^2 \right)}_{O_p \left( \frac{1}{q} \right)} + \left( \kappa \sigma_c^4 + 3 (\sigma_c^2 + \sigma_z^2)^2 \right) + O_p \left( \frac{1}{q} \right) \\
&= \kappa \sigma_c^4 + 3 (\sigma_c^2 + \sigma_z^2)^2 + O_p \left( \frac{1}{q} \right).
\end{aligned} \tag{14}$$

It follows that  $\bar{r}^{(4)} \xrightarrow{p} \kappa \sigma_c^4 + 3 (\sigma_c^2 + \sigma_z^2)^2$  as  $q \rightarrow \infty$ , i.e. as  $n, p \rightarrow \infty$ .

Plugging in the excess kurtosis,  $\kappa = 3$ , of Laplace distributed  $c_{ij}$ , combining  $\bar{r}^{(4)} \xrightarrow{p} 3\sigma_c^4 + 3(\sigma_c^2 + \sigma_z^2)^2$  as  $n, p \rightarrow \infty$  with  $\bar{r}^{(2)} \xrightarrow{p} \sigma_c^2 + \sigma_z^2$  as  $n, p \rightarrow \infty$  and applying the continuous mapping theorem yields  $\hat{\sigma}_c^4 \xrightarrow{p} \sigma_c^4$ ,  $\hat{\sigma}_c^2 \xrightarrow{p} \sigma_c^2$ ,  $\hat{\lambda}_c \xrightarrow{p} \lambda_c$  and  $\hat{\sigma}_z^2 \xrightarrow{p} \sigma_z^2$  as  $n, p \rightarrow \infty$ .

**Derivation of Asymptotic Variance of  $\hat{\sigma}_c^2$ :** Letting  $x_{ij} = c_{ij} + z_{ij}$ , we will use the following decomposition:

$$r_{ij} = x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}$$

Without loss of generality, we let  $n = \pi_1 q$ ,  $p = \pi_2 q$ , where  $0 < \pi_1 < 1$  and  $0 < \pi_2 < 1$  are fixed and  $\pi_1 + \pi_2 = 1$ . Then the scenario  $n, p \rightarrow \infty$  is equivalent to  $q \rightarrow \infty$ .

For  $\bar{r}^{(2)}$  we have:

$$\begin{aligned}
\bar{r}^{(2)} &= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}_{\cdot\cdot})^2 \\
&= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \left( x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + O_p\left(\frac{1}{q}\right) \right)^2 \\
&= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j})^2 + (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j}) O_p\left(\frac{1}{q}\right) + O_p\left(\frac{1}{q^2}\right) \\
&= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j})^2 + O_p\left(\frac{1}{q^2}\right) \\
&= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 + \bar{x}_{i\cdot}^2 + \bar{x}_{\cdot j}^2 - 2x_{ij}\bar{x}_{i\cdot} - 2x_{ij}\bar{x}_{\cdot j} + 2\bar{x}_{i\cdot}\bar{x}_{\cdot j} + O_p\left(\frac{1}{q^2}\right)
\end{aligned}$$

We find the order of most of the terms below:

$$\begin{aligned}
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij} \bar{x}_{i\cdot} &= \left( \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij} \right) O_p\left(\frac{1}{\sqrt{q}}\right) \\
&= O_p\left(\frac{1}{\sqrt{q^3}}\right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij} \bar{x}_{\cdot j} &= O_p\left(\frac{1}{\sqrt{q^3}}\right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \bar{x}_{i\cdot} \bar{x}_{\cdot j} &= \left( \frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \bar{x}_{i\cdot} \right) \left( \frac{1}{\pi_2 q} \sum_{j=1}^{\pi_2 q} \bar{x}_{\cdot j} \right) \\
&= \bar{x}_{\cdot\cdot}^2 \\
&= O_p\left(\frac{1}{q^2}\right).
\end{aligned}$$

Plugging these expressions into the equation for  $\bar{r}^{(2)}$  gives:

$$\bar{r}^{(2)} = \bar{x}^{(2)} + \frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \bar{x}_{i\cdot}^2 + \frac{1}{\pi_2 q} \sum_{j=1}^{\pi_2 q} \bar{x}_{\cdot j}^2 + O_p\left(\frac{1}{\sqrt{q^3}}\right).$$

Now we compute the means and find the orders of the variances of the middle two terms,

starting with  $\frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \bar{x}_{i\cdot}^2$ .

$$\begin{aligned}
\mathbb{E} \left[ \frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \bar{x}_{i\cdot}^2 \right] &= \mathbb{E} [\bar{x}_{i\cdot}^2] \\
&= \frac{1}{\pi_2^2 q^2} \mathbb{E} \left[ \left( \sum_{j=1}^{\pi_2 q} x_{ij} \right)^2 \right] \\
&= \frac{1}{\pi_2^2 q^2} (\pi_2 q \mathbb{E} [x_{ij}^2]) \\
&= \frac{\mathbb{E} [x_{ij}^2]}{\pi_2 q}.
\end{aligned}$$

Note that  $\frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \bar{x}_{i\cdot}^2$  is a mean over independent variables,  $\bar{x}_{i\cdot}^2$ . Accordingly,  $\mathbb{V} \left[ \frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \bar{x}_{i\cdot}^2 \right] = \frac{1}{\pi_1 q} \mathbb{V} [\bar{x}_{i\cdot}^2]$ . Now we find the order of  $\mathbb{V} [\bar{x}_{i\cdot}^2]$ :

$$\begin{aligned}
\mathbb{V} [\bar{x}_{i\cdot}^2] &= \mathbb{E} [\bar{x}_{i\cdot}^4] - \mathbb{E} [\bar{x}_{i\cdot}^2]^2 \\
&= \frac{1}{\pi_2^4 q^4} \mathbb{E} \left[ \left( \sum_{j=1}^{\pi_2 q} x_{ij} \right)^4 \right] + O \left( \frac{1}{q^2} \right) \\
&= \frac{1}{\pi_2^4 q^4} \left( \pi_2 q \mathbb{E} [x_{ij}^4] + 3\pi_2 q (\pi_2 q - 1) \mathbb{E} [x_{ij}^2]^2 \right) + O \left( \frac{1}{q^2} \right) \\
&= O \left( \frac{1}{q^2} \right).
\end{aligned}$$

Then  $\mathbb{V} \left[ \frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \bar{x}_{i\cdot}^2 \right] = O \left( \frac{1}{q^3} \right)$  and  $\frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \bar{x}_{i\cdot}^2 - \frac{\mathbb{E} [x_{ij}^2]}{\pi_2 q} = O_p \left( \frac{1}{\sqrt{q^3}} \right)$ . The same logic yields  $\frac{1}{\pi_2 q} \sum_{j=1}^{\pi_2 q} \bar{x}_{\cdot j}^2 - \frac{\mathbb{E} [x_{ij}^2]}{\pi_1 q} = O_p \left( \frac{1}{\sqrt{q^3}} \right)$ . Plugging this into the equation for  $\bar{r}^{(2)}$  gives:

$$\begin{aligned}
\bar{r}^{(2)} &= \bar{x}^{(2)} + \frac{\mathbb{E} [x_{ij}^2]}{\pi_1 q} + \frac{\mathbb{E} [x_{ij}^2]}{\pi_2 q} + O_p \left( \frac{1}{\sqrt{q^3}} \right) \\
&= \bar{x}^{(2)} + \frac{\mathbb{E} [x_{ij}^2]}{\pi_1 \pi_2 q} + O_p \left( \frac{1}{\sqrt{q^3}} \right)
\end{aligned}$$

For  $\bar{r}^{(4)}$  we have:



$$\begin{aligned}
\bar{r}^{(4)} &= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^4 \\
&= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \left( x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + O_p \left( \frac{1}{q} \right) \right)^4 \\
&= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j})^4 + (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j})^3 O_p \left( \frac{1}{q} \right) + \\
&\quad (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j})^2 O_p \left( \frac{1}{q^2} \right) + (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j}) O_p \left( \frac{1}{q^3} \right) + O_p \left( \frac{1}{q^4} \right) \\
&= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j})^4 + O_p \left( \frac{1}{q^2} \right) + O_p \left( \frac{1}{q^2} \right) + O_p \left( \frac{1}{q^4} \right) + O_p \left( \frac{1}{q^4} \right) \\
&= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j})^4 + O_p \left( \frac{1}{q^2} \right) \\
&= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^4 + \bar{x}_{i.}^4 + \bar{x}_{.j}^4 + 4x_{ij}^3 \bar{x}_{i.} + 4x_{ij} \bar{x}_{i.}^3 + 4x_{ij}^3 \bar{x}_{.j} + 4x_{ij} \bar{x}_{.j}^3 + 4\bar{x}_{i.}^3 \bar{x}_{.j} + 4\bar{x}_{i.} \bar{x}_{.j}^3 + \\
&\quad 6x_{ij}^2 \bar{x}_{i.}^2 + 6x_{ij}^2 \bar{x}_{.j}^2 + 6\bar{x}_{i.}^2 \bar{x}_{.j}^2 + 12x_{ij}^2 \bar{x}_{i.} \bar{x}_{.j} + 12x_{ij} \bar{x}_{i.}^2 \bar{x}_{.j} + 12x_{ij} \bar{x}_{i.} \bar{x}_{.j}^2 + O_p \left( \frac{1}{q^2} \right)
\end{aligned}$$

As with  $\bar{r}^{(2)}$ , we can find the order of most of these terms:

$$\begin{aligned}
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \bar{x}_i^4 &= O_p \left( \frac{1}{q^2} \right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \bar{x}_{\cdot j}^4 &= O_p \left( \frac{1}{q^2} \right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^3 \bar{x}_i &= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^3 O_p \left( \frac{1}{\sqrt{q}} \right) = O_p \left( \frac{1}{\sqrt{q}^3} \right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^3 \bar{x}_{\cdot j} &= O_p \left( \frac{1}{\sqrt{q}^3} \right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij} \bar{x}_i^3 &= \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij} O_p \left( \frac{1}{\sqrt{q}^3} \right) = O_p \left( \frac{1}{\sqrt{q}^5} \right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij} \bar{x}_{\cdot j}^3 &= O_p \left( \frac{1}{\sqrt{q}^5} \right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \bar{x}_i \bar{x}_{\cdot j}^3 &= O_p \left( \frac{1}{\sqrt{q}} \right) O_p \left( \frac{1}{\sqrt{q}^3} \right) = O_p \left( \frac{1}{q^2} \right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \bar{x}_i^3 \bar{x}_{\cdot j} &= O_p \left( \frac{1}{q^2} \right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \bar{x}_i^2 \bar{x}_{\cdot j}^2 &= O_p \left( \frac{1}{q} \right) O_p \left( \frac{1}{q} \right) = O_p \left( \frac{1}{q^2} \right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij} \bar{x}_i^2 \bar{x}_{\cdot j} &= O_p \left( \frac{1}{q} \right) O_p \left( \frac{1}{q} \right) O_p \left( \frac{1}{\sqrt{q}} \right) = O_p \left( \frac{1}{\sqrt{q}^5} \right) \\
\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij} \bar{x}_i \bar{x}_{\cdot j}^2 &= O_p \left( \frac{1}{\sqrt{q}^5} \right).
\end{aligned}$$

Plugging these expressions into the equation for  $\bar{r}^{(2)}$  gives:

$$\bar{r}^{(4)} = \bar{x}^{(4)} + \frac{6}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i^2 + \frac{6}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_{\cdot j}^2 + \frac{12}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i \bar{x}_{\cdot j} + O_p \left( \frac{1}{\sqrt{q}^3} \right).$$

Now we compute the means and find the orders of the variances of the remaining terms, starting with  $\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i^2$ :

$$\begin{aligned}
\mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i^2 \right] &= \frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \mathbb{E} \left[ \bar{x}_i^2 \left( \frac{1}{\pi_2 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \right) \right] \\
&= \frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \mathbb{E} \left[ \left( \frac{1}{\pi_2 q} \sum_{j=1}^{\pi_2 q} x_{ij} \right)^2 \left( \frac{1}{\pi_2 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \right) \right] \\
&= \frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \left( \frac{1}{\pi_2^3 q^3} \mathbb{E} \left[ \left( \sum_{j=1}^{\pi_2 q} x_{ij} \right)^2 \left( \sum_{j=1}^{\pi_2 q} x_{ij}^2 \right) \right] \right) \\
&= \frac{1}{\pi_1 q} \sum_{i=1}^{\pi_1 q} \left( \frac{1}{\pi_2^3 q^3} \left( \pi_2 q \mathbb{E} [x_{ij}^4] + \pi_2 q (\pi_2 q - 1) \mathbb{E} [x_{ij}^2]^2 \right) \right) \\
&= \frac{\mathbb{E} [x_{ij}^4] + (\pi_2 q - 1) \mathbb{E} [x_{ij}^2]^2}{\pi_2^2 q^2} \\
&= O \left( \frac{1}{q^2} \right) + \frac{\mathbb{E} [x_{ij}^2]^2}{\pi_2 q}.
\end{aligned}$$

Note that this term is the mean of  $\pi_1 q$  independent, identically distributed terms, so to find order of its variance we just need to find the order of the variance of a single one of the terms.

$$\begin{aligned}
\mathbb{V} \left[ \bar{x}_i^2 \left( \frac{1}{\pi_2 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \right) \right] &= \mathbb{E} \left[ \left( \bar{x}_i^2 \left( \frac{1}{\pi_2 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \right) \right)^2 \right] - \mathbb{E} \left[ \bar{x}_i^2 \left( \frac{1}{\pi_2 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \right) \right]^2 \\
&= \frac{1}{\pi_2^6 q^6} \mathbb{E} \left[ \left( \sum_{j=1}^{\pi_2 q} x_{ij} \right)^4 \left( \sum_{j=1}^{\pi_2 q} x_{ij}^2 \right)^2 \right] + O \left( \frac{1}{q^2} \right) \\
&= \frac{1}{\pi_2^6 q^6} \left( \pi_2 q \mathbb{E} [x_{ij}^8] + \pi_2 q (\pi_2 q - 1) \mathbb{E} [x_{ij}^4]^2 + 2 \pi_2 q (\pi_2 q - 1) \mathbb{E} [x_{ij}^6] \mathbb{E} [x_{ij}^2] + \right. \\
&\quad \left. 2 \pi_2 q (\pi_2 q - 1) (\pi_2 q - 2) \mathbb{E} [x_{ij}^4] \mathbb{E} [x_{ij}^2]^2 + 6 \pi_2 q (\pi_2 q - 1) \mathbb{E} [x_{ij}^6] \mathbb{E} [x_{ij}^2] + \right. \\
&\quad \left. 12 \pi_2 q (\pi_2 q - 1) \mathbb{E} [x_{ij}^4] \mathbb{E} [x_{ij}^4] + 12 \pi_2 q (\pi_2 q - 1) (\pi_2 q - 2) (\pi_2 q - 3) \mathbb{E} [x_{ij}^2]^4 \right) + \\
&\quad O \left( \frac{1}{q^2} \right) = O \left( \frac{1}{q^2} \right).
\end{aligned}$$

It follows that  $\mathbb{V} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i^2 \right] = O \left( \frac{1}{q^3} \right)$  and

$$\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i^2 - \mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i^2 \right] = O_p \left( \frac{1}{\sqrt{q^3}} \right).$$

By the same logic,  $\mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_{\cdot j} \right] = O \left( \frac{1}{q^2} \right) + \frac{\mathbb{E}[x_{ij}^2]^2}{\pi_1 q}$  and

$$\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_{\cdot j}^2 - \mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_{\cdot j}^2 \right] = O_p \left( \frac{1}{\sqrt{q^3}} \right).$$

Now we just have to compute the mean and order of the variance of one last term,  $\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}$ .

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j} \right] &= \left( \frac{1}{\pi_1^2 \pi_2^2 q^4} \right) \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \mathbb{E} \left[ x_{ij}^2 \left( \sum_{j'=1}^{\pi_2 q} x_{ij'} \right) \left( \sum_{i'=1}^{\pi_1 q} x_{i'j} \right) \right] \\ &= \frac{\pi_1 \pi_2 q^2 \mathbb{E}[x_{ij}^4]}{\pi_1^2 \pi_2^2 q^4} = \frac{\mathbb{E}[x_{ij}^4]}{\pi_1 \pi_2 q^2} = O \left( \frac{1}{q^2} \right). \end{aligned}$$

Because this term cannot be written as an average over independent, identically distributed random variables, computing the variance in this case is a little bit trickier. We have

$$\begin{aligned} \mathbb{V} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j} \right] &= \frac{1}{\pi_1^2 \pi_2^2 q^4} \left( \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \mathbb{V} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}] + 2 \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \sum_{j'=1, j' \neq j}^{\pi_2 q} \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{ij'}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j'}] + \right. \\ &\quad \left. 2 \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \sum_{i'=1, i' \neq i}^{\pi_1 q} \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{i'j}^2 \bar{x}_{i' \cdot} \bar{x}_{\cdot j}] + \right. \\ &\quad \left. 2 \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} \sum_{i'=1, i' \neq i}^{\pi_1 q} \sum_{j'=1, j' \neq j}^{\pi_2 q} \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{i'j'}^2 \bar{x}_{i' \cdot} \bar{x}_{\cdot j'}] \right) \\ &= \frac{1}{\pi_1^2 \pi_2^2 q^4} \left( \pi_1 \pi_2 q^2 \mathbb{V} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}] + 2 \pi_1 \pi_2 q^2 (\pi_2 q - 1) \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{ij'}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j'}] + \right. \\ &\quad \left. 2 \pi_1 \pi_2 q^2 (\pi_2 q - 1) \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{i'j}^2 \bar{x}_{i' \cdot} \bar{x}_{\cdot j}] + \right. \\ &\quad \left. 2 \pi_1 \pi_2 q^2 (\pi_1 \pi_2 q^2 - \pi_1 q - \pi_2 q + 1) \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{i'j'}^2 \bar{x}_{i' \cdot} \bar{x}_{\cdot j'}] \right) \\ &= \frac{1}{\pi_1 \pi_2 q^2} \mathbb{V} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}] + \frac{2 (\pi_2 q - 1)}{\pi_1 \pi_2 q^2} \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{ij'}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j'}] + \\ &\quad \frac{2 (\pi_2 q - 1)}{\pi_1 \pi_2 q^2} \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{i'j}^2 \bar{x}_{i' \cdot} \bar{x}_{\cdot j}] + \\ &\quad \frac{2 (\pi_1 \pi_2 q^2 - \pi_1 q - \pi_2 q + 1)}{\pi_1 \pi_2 q^2} \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{i'j'}^2 \bar{x}_{i' \cdot} \bar{x}_{\cdot j'}] \\ &= O \left( \frac{1}{q^2} \right) \mathbb{V} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}] + O \left( \frac{1}{q} \right) \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{ij'}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j'}] + \\ &\quad O \left( \frac{1}{q} \right) \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{i'j}^2 \bar{x}_{i' \cdot} \bar{x}_{\cdot j}] + O(1) \text{Cov} [x_{ij}^2 \bar{x}_{i \cdot} \bar{x}_{\cdot j}, x_{i'j'}^2 \bar{x}_{i' \cdot} \bar{x}_{\cdot j'}]. \end{aligned}$$

Examining the first term we have:

$$\begin{aligned}
\mathbb{V} [x_{ij}^2 \bar{x}_i \bar{x}_j] &= \mathbb{E} [x_{ij}^4 \bar{x}_i^2 \bar{x}_j^2] - \mathbb{E} [x_{ij}^2 \bar{x}_i \bar{x}_j]^2 \\
&= \left( \frac{1}{\pi_1^2 \pi_2^2 q^4} \right) \mathbb{E} \left[ x_{ij}^4 \left( \sum_{j'=1}^{\pi_2 q} x_{ij'} \right)^2 \left( \sum_{i'=1}^{\pi_1 q} x_{i'j} \right)^2 \right] + O \left( \frac{1}{q^4} \right) \\
&= \left( \frac{\mathbb{E} [x_{ij}^8] + \pi_1 q (\pi_2 q - 1) \mathbb{E} [x_{ij}^4] \mathbb{E} [x_{ij}^2]^2}{\pi_1^2 \pi_2^2 q^4} \right) + O \left( \frac{1}{q^4} \right) = O \left( \frac{1}{q^2} \right).
\end{aligned}$$

Then examining the second term we have:

$$\begin{aligned}
\text{Cov} [x_{ij}^2 \bar{x}_i \bar{x}_j, x_{ij'}^2 \bar{x}_i \bar{x}_{j'}] &= \mathbb{E} [x_{ij}^2 x_{ij'}^2 \bar{x}_i^2 \bar{x}_j \bar{x}_{j'}] - \mathbb{E} [x_{ij}^2 \bar{x}_i \bar{x}_j]^2 \\
&= \frac{1}{\pi_1^2 \pi_2^2 q^4} \mathbb{E} \left[ x_{ij}^2 x_{ij'}^2 \left( \sum_{j''=1}^{\pi_2 q} x_{ij''} \right)^2 \left( \sum_{i'=1}^{\pi_1 q} x_{i'j} \right) \left( \sum_{i''=1}^{\pi_1 q} x_{i''j} \right) \right] + O \left( \frac{1}{q^4} \right) \\
&= \frac{2 \mathbb{E} [x_{ij}^4 x_{ij'}^4]}{\pi_1^2 \pi_2^2 q^4} + O \left( \frac{1}{q^4} \right) = O \left( \frac{1}{q^4} \right).
\end{aligned}$$

The same logic yields  $\text{Cov} [x_{ij}^2 \bar{x}_i \bar{x}_j, x_{ij'}^2 \bar{x}_i \bar{x}_{j'}] = O \left( \frac{1}{q^4} \right)$ .

Then examining the last term yields:

$$\begin{aligned}
\text{Cov} [x_{ij}^2 \bar{x}_i \bar{x}_j, x_{i'j'}^2 \bar{x}_{i'} \bar{x}_{j'}] &= \mathbb{E} [x_{ij}^2 x_{i'j'}^2 \bar{x}_i \bar{x}_{i'} \bar{x}_j \bar{x}_{j'}] - \mathbb{E} [x_{ij}^2 \bar{x}_i \bar{x}_j]^2 \\
&= \mathbb{E} [x_{ij}^2 \bar{x}_i \bar{x}_j] \mathbb{E} [x_{i'j'}^2 \bar{x}_{i'} \bar{x}_{j'}] - \mathbb{E} [x_{ij}^2 \bar{x}_i \bar{x}_j]^2 = 0.
\end{aligned}$$

Putting the pieces together, we get:

$$\mathbb{V} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i \bar{x}_j \right] = O \left( \frac{1}{q^2} \right) O \left( \frac{1}{q^2} \right) + O \left( \frac{1}{q} \right) O \left( \frac{1}{q^4} \right) + O \left( \frac{1}{q} \right) O \left( \frac{1}{q^4} \right) = O \left( \frac{1}{q^4} \right).$$

It follows that

$$\frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i \bar{x}_j - \mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i \bar{x}_j \right] = O_p \left( \frac{1}{q^2} \right).$$

Substituting these expressions into our last equation for  $\bar{r}^{(4)}$ , we get:

$$\begin{aligned}
\bar{r}^{(4)} &= \bar{x}^{(4)} + 6 \left( \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i^2 - \mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i^2 \right] \right) + \mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i^2 \right] + \\
&6 \left( \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_{\cdot j}^2 - \mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_{\cdot j}^2 \right] \right) + 6 \mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_{\cdot j}^2 \right] + \\
&12 \left( \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i \bar{x}_{\cdot j} - \mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i \bar{x}_{\cdot j} \right] \right) + 12 \mathbb{E} \left[ \frac{1}{\pi_1 \pi_2 q^2} \sum_{i=1}^{\pi_1 q} \sum_{j=1}^{\pi_2 q} x_{ij}^2 \bar{x}_i \bar{x}_{\cdot j} \right] + \\
&O_p \left( \frac{1}{\sqrt{q}^3} \right) \\
&= \bar{x}^{(4)} + O_p \left( \frac{1}{\sqrt{q}^3} \right) + O \left( \frac{1}{q^2} \right) + \frac{6 \mathbb{E} [x_{ij}^2]^2}{\pi_1 q} + \\
&O_p \left( \frac{1}{\sqrt{q}^3} \right) + O \left( \frac{1}{q^2} \right) + \frac{6 \mathbb{E} [x_{ij}^2]^2}{\pi_2 q} + O_p \left( \frac{1}{q^2} \right) + O \left( \frac{1}{q^2} \right) + O_p \left( \frac{1}{\sqrt{q}^3} \right) \\
&= \bar{x}^{(4)} + \frac{6 \mathbb{E} [x_{ij}^2]^2}{\pi_1 \pi_2 q} + O \left( \frac{1}{q^2} \right) + O_p \left( \frac{1}{\sqrt{q}^3} \right).
\end{aligned}$$

Now we plug these expressions for  $\bar{r}^{(2)}$  and  $\bar{r}^{(4)}$  into our equation for  $\hat{\sigma}_c^4$ :

$$\begin{aligned}
\hat{\sigma}_c^4 &= O(1) \left( \bar{r}^{(4)}/3 - (\bar{r}^{(2)})^2 \right) \\
&= O(1) \left( \bar{x}^{(4)}/3 + \frac{2\mathbb{E}[x_{ij}^2]^2}{\pi_1\pi_2q} + O\left(\frac{1}{q^2}\right) + O_p\left(\frac{1}{\sqrt{q^3}}\right) - \left( \bar{x}^{(2)} + \frac{\mathbb{E}[x_{ij}^2]}{\pi_1\pi_2q} + O_p\left(\frac{1}{\sqrt{q^3}}\right) \right)^2 \right) \\
&= O(1) \left( \bar{x}^{(4)}/3 - (\bar{x}^{(2)})^2 + \frac{2\mathbb{E}[x_{ij}^2]^2}{\pi_1\pi_2q} + O\left(\frac{1}{q^2}\right) + O_p\left(\frac{1}{\sqrt{q^3}}\right) - \right. \\
&\quad \left. \underbrace{\left( \frac{\mathbb{E}[x_{ij}^2]}{\pi_1\pi_2q} \right)^2}_{O\left(\frac{1}{q^2}\right)} + O_p\left(\frac{1}{q^3}\right) + O_p\left(\frac{1}{\sqrt{q^3}}\right) + O_p\left(\frac{1}{\sqrt{q^5}}\right) - 2(\bar{x}^{(2)}) \left( \frac{\mathbb{E}[x_{ij}^2]}{\pi_1\pi_2q} \right) \right) \\
&= O(1) \left( \bar{x}^{(4)}/3 - (\bar{x}^{(2)})^2 + \frac{2\mathbb{E}[x_{ij}^2]^2}{\pi_1\pi_2q} - 2 \underbrace{(\bar{x}^{(2)} - \mathbb{E}[x_{ij}^2])}_{O_p\left(\frac{1}{q}\right)} \underbrace{\left( \frac{\mathbb{E}[x_{ij}^2]}{\pi_1\pi_2q} \right)}_{O\left(\frac{1}{q}\right)} - 2 \left( \frac{\mathbb{E}[x_{ij}^2]^2}{\pi_1\pi_2q} \right) + \right. \\
&\quad \left. O\left(\frac{1}{q^2}\right) + O_p\left(\frac{1}{\sqrt{q^3}}\right) \right) \\
&= O(1) \left( \bar{x}^{(4)}/3 - (\bar{x}^{(2)})^2 + O\left(\frac{1}{q^2}\right) + O_p\left(\frac{1}{\sqrt{q^3}}\right) \right).
\end{aligned}$$

Recalling that  $\kappa$  is defined such that  $\mathbb{E}[x_{ij}^4]/3 - \mathbb{E}[x_{ij}^2]^2 = \kappa\sigma_c^4/3$ , we have

$$\sqrt{\pi_1\pi_2q^2} (\hat{\sigma}_c^4 - \kappa\sigma_c^4/3) \xrightarrow{d} \sqrt{\pi_1\pi_2q^2} (\bar{x}^{(4)}/3 - (\bar{x}^{(2)})^2 - \kappa\sigma_c^4/3)$$

as  $q \rightarrow \infty$ .

But what is the distribution of the right hand side? As  $\bar{x}^{(2)}$  and  $\bar{x}^{(4)}$  are sample moments of independent, identically distributed random variables  $x_{ij}$ , we know that as  $q \rightarrow \infty$ :

$$\sqrt{\pi_1\pi_2q^2} \begin{pmatrix} \bar{x}^{(2)} - \mathbb{E}[x_{ij}^2] \\ \bar{x}^{(4)} - \mathbb{E}[x_{ij}^4] \end{pmatrix} \xrightarrow{d} N \left( \mathbf{0}, \begin{pmatrix} \mathbb{V}[x_{ij}^2] & \text{Cov}[x_{ij}^2, x_{ij}^4] \\ \text{Cov}[x_{ij}^2, x_{ij}^4] & \mathbb{V}[x_{ij}^4] \end{pmatrix} \right).$$

Let  $f(y, z) = z/3 - y^2$  and note that  $f(\mathbb{E}[x_{ij}^2], \mathbb{E}[x_{ij}^4]) = \kappa\sigma_c^4/3$  and

$$\nabla f(\mathbb{E}[x_{ij}^2], \mathbb{E}[x_{ij}^4]) = \begin{pmatrix} -2\mathbb{E}[x_{ij}^2] \\ 1/3 \end{pmatrix}.$$

Applying the delta method yields:

$$\sqrt{\pi_1 \pi_2 q^2} \left( \bar{x}^{(4)} / 3 - (\bar{x}^{(2)})^2 - \kappa \sigma_c^4 / 3 \right) \xrightarrow{d} N \left( 0, 4\mathbb{E}[x_{ij}^2]^2 \mathbb{V}[x_{ij}^2] - 4\mathbb{E}[x_{ij}^2] \text{Cov}[x_{ij}^2, x_{ij}^4] / 3 + \mathbb{V}[x_{ij}^4] / 9 \right) \quad (15)$$

as  $q \rightarrow \infty$ .

Up to this point, we have just used the assumption that  $x_{ij} = c_{ij} + z_{ij}$ , where  $z_{ij}$  is normal and both  $c_{ij}$  and  $z_{ij}$  are independent, identically distributed from symmetric, mean zero distributions. This will be convenient later, when we need to find the asymptotic distribution of  $\hat{\sigma}_c^4$  assuming (a)  $c_{ij}$  are normally distributed and/or exactly equal to zero and (b) assuming  $c_{ij}$  are drawn from a Bernoulli-normal spike-and-slab prior. However here, we are interested in finding the asymptotic variance of  $\hat{\sigma}_c^4$  assuming  $c_{ij}$  are Laplace distributed with variance  $\sigma_c^2$ . In this case, we have:

$$\begin{aligned} \mathbb{E}[x_{ij}^2] &= \sigma_c^2 + \sigma_z^2 \\ \mathbb{E}[x_{ij}^4] &= 3\sigma_c^4 + 3(\sigma_c^2 + \sigma_z^2)^2 \\ \mathbb{V}[x_{ij}^2] &= 5\sigma_c^4 + 4\sigma_c^2\sigma_z^2 + 2\sigma_z^4 \\ \text{Cov}[x_{ij}^2, x_{ij}^4] &= 6(14\sigma_c^6 + 13\sigma_c^4\sigma_z^2 + 6\sigma_c^2\sigma_z^4 + 2\sigma_z^6) \\ \mathbb{V}[x_{ij}^4] &= 12(207\sigma_c^8 + 204\sigma_c^6\sigma_z^2 + 99\sigma_c^4\sigma_z^4 + 32\sigma_c^2\sigma_z^6 + 8\sigma_z^8). \end{aligned}$$

Substituting these in to our expression for the asymptotic variance of  $\hat{\sigma}_c^4$  and simplifying yields and substituting  $n$  and  $p$  back in for  $\pi_1 q$  and  $\pi_2 q$ , we have

$$\sqrt{np} (\hat{\sigma}_c^4 - \sigma_c^4) \xrightarrow{d} N(0, 8(69\sigma_c^8 + 42\sigma_c^6\sigma_z^2 + 15\sigma_c^4\sigma_z^4 + 4\sigma_c^2\sigma_z^6 + \sigma_z^8)/3) \quad (16)$$

as  $n, p \rightarrow \infty$ .

**Proof of Proposition 3.1:** The proof of Proposition 3.1 follows directly from the derivation of the asymptotic distribution of  $\hat{\sigma}_c^4$  given in the previous section. To prove Proposition 3.1, we need to find the asymptotic distribution of  $\hat{\sigma}_c^4$  under the null hypothesis,  $H_0$ :  $c_{ij} + z_{ij} \sim \text{i.i.d. } N(0, \sigma_c^2 + \sigma_z^2)$ . Under this null hypothesis,  $c_{ij} + z_{ij} \stackrel{d}{=} \tilde{c}_{ij} + \tilde{z}_{ij}$ , where  $\tilde{c}_{ij}$  is Laplace distributed with 0 and  $\tilde{z}_{ij}$  is normally distributed with variance  $\sigma_c^2 + \sigma_z^2$ . Applying equation (16), yields:

$$\sqrt{np} \hat{\sigma}_c^4 \xrightarrow{d} N \left( 0, 8(\sigma_c^2 + \sigma_z^2)^4 / 3 \right)$$

as  $n, p \rightarrow \infty$ .

However when we are testing the null hypothesis,  $\sigma_c^2 + \sigma_z^2$  is unknown. Fortunately,  $(\hat{\sigma}_c^2 + \hat{\sigma}_z^2)^4 = \left( \left( \frac{np}{(n-1)(p-1)} \right) \bar{r}^{(2)} \right)^4 \xrightarrow{p} \sigma_c^2 + \sigma_z^2$  as  $n, p \rightarrow \infty$ .

It follows directly from Equations (17) and (13) that  $\hat{\sigma}_c^2 + \hat{\sigma}_z^2 \xrightarrow{p} \sigma_c^2 + \sigma_z^2$  as  $n, p \rightarrow \infty$ , and



by the continuous mapping theorem  $(\hat{\sigma}_c^2 + \hat{\sigma}_z^2)^4 \xrightarrow{p} (\sigma_c^2 + \sigma_z^2)^4$  as  $n, p \rightarrow \infty$ . Accordingly,

$$\sqrt{np} \left( \frac{\hat{\sigma}_c^4}{\sqrt{\frac{8}{3}} (\hat{\sigma}_c^2 + \hat{\sigma}_z^2)^2} \right) \xrightarrow{d} N(0, 1)$$

as  $n, p \rightarrow \infty$  and under  $H_0$ ,  $\Pr \left( \sqrt{np} \left( \frac{\hat{\sigma}_c^4}{\sqrt{\frac{8}{3}} (\hat{\sigma}_c^2 + \hat{\sigma}_z^2)^2} \right) > z_{1-\alpha} \right) \rightarrow \alpha$  as  $n, p \rightarrow \infty$ .

**Proof of Proposition 3.2:** Combining Equation (16) and consistency of  $\hat{\sigma}_c^2 + \hat{\sigma}_z^2$  for  $\sigma_c^2 + \sigma_z^2$  as  $n, p \rightarrow \infty$  that follows from Equations (17) and (13), we have:

$$\sqrt{np} \left( \frac{\hat{\sigma}_c^4}{\sqrt{\frac{8}{3}} (\hat{\sigma}_c^2 + \hat{\sigma}_z^2)^2} - \frac{\sigma_c^4}{\sqrt{\frac{8}{3}} (\sigma_c^2 + \sigma_z^2)^2} \right) \xrightarrow{d} N \left( 0, \frac{69\sigma_c^8 + 42\sigma_c^6\sigma_z^2 + 15\sigma_c^4\sigma^4 + 4\sigma_c^2\sigma_z^6 + \sigma_z^8}{(\sigma_c^4 + \sigma_z^2)^4} \right).$$

We can simplify the mean and variance and write them as a function of  $\phi^2 = \sigma_c^2/\sigma_z^2$ :

$$\begin{aligned} \frac{\sigma_c^4}{\sqrt{\frac{8}{3}} (\sigma_c^2 + \sigma_z^2)^2} &= \sqrt{\frac{3}{8}} \left( \frac{\phi^4}{(\phi^2 + 1)^2} \right) \\ \frac{69\sigma_c^8 + 42\sigma_c^6\sigma_z^2 + 15\sigma_c^4\sigma^4 + 4\sigma_c^2\sigma_z^6 + \sigma_z^8}{(\sigma_c^4 + \sigma_z^2)^4} &= \frac{68\sigma_c^8 + 36\sigma_c^6\sigma_z^2 + 9\sigma_c^4\sigma^4 + (\sigma_c^2 + \sigma_z^2)^4}{(\sigma_c^4 + \sigma_z^2)^4} \\ &= 1 + \frac{68\sigma_c^8 + 36\sigma_c^6\sigma_z^2 + 9\sigma_c^4\sigma^4}{(\sigma_c^4 + \sigma_z^2)^4} \\ &= 1 + \frac{68\phi^8 + 36\phi^6 + 9\phi^4}{(\phi^4 + 1)^4}. \end{aligned}$$

It follows that:

$$\Pr \left( \sqrt{np} \left( \frac{\hat{\sigma}_c^4}{\sqrt{\frac{8}{3}} (\hat{\sigma}_c^2 + \hat{\sigma}_z^2)^2} \right) > z_{1-\alpha} \right) \rightarrow 1 - \Phi \left( \frac{z_{1-\alpha} - \sqrt{\frac{3np}{8}} \left( \frac{\phi^4}{(\phi^2+1)^2} \right)}{\sqrt{1 + \frac{68\phi^8 + 36\phi^6 + 9\phi^4}{(\phi^4+1)^4}}} \right)$$

as  $n, p \rightarrow \infty$ .

**Proof of Proposition 3.3:** To prove Proposition 3.3, we first need to compute the unknown quantities that appear in Equation (15):

$$\begin{aligned}
\mathbb{E}[x_{ij}^2] &= \pi_c \tau_c^2 + \sigma_z^2 \\
\mathbb{E}[x_{ij}^4] &= 3(\pi_c \tau_c^4 + 2\pi_c \tau_c^2 \sigma_z^2 + \sigma_z^4) \\
\mathbb{E}[x_{ij}^6] &= \mathbb{E}[c_{ij}^6] + 15\mathbb{E}[c_{ij}^4] \mathbb{E}[z_{ij}^2] + 15\mathbb{E}[c_{ij}^2] \mathbb{E}[z_{ij}^4] + \mathbb{E}[z_{ij}^6] \\
&= 15(\pi_c \tau_c^6 + 3\pi_c \tau_c^4 \sigma_z^2 + 3\pi_c \tau_c^2 \sigma_z^4 + \sigma_z^6) \\
\mathbb{E}[x_{ij}^8] &= \mathbb{E}[c_{ij}^8] + 28\mathbb{E}[c_{ij}^6] \mathbb{E}[z_{ij}^2] + 70\mathbb{E}[c_{ij}^4] \mathbb{E}[z_{ij}^4] + 28\mathbb{E}[c_{ij}^2] \mathbb{E}[z_{ij}^6] + \mathbb{E}[z_{ij}^8] \\
&= 105(\pi_c \tau_c^8 + 4\pi_c \tau_c^6 \sigma_z^2 + 6\pi_c \tau_c^4 \sigma_z^4 + 4\pi_c \tau_c^2 \sigma_z^6 + \sigma_z^8) \\
\mathbb{V}[x_{ij}^2] &= (3 - \pi_c) \pi_c \tau_c^4 + 4\pi_c \tau_c^2 \sigma_z^2 + 2\sigma_z^2 \\
\text{Cov}[x_{ij}^2, x_{ij}^4] &= 12\sigma_z^6 + 36\pi_c \sigma_z^4 \tau_c^2 + 42\pi_c \sigma_z^2 \tau_c^4 - 6\pi_c^2 \sigma_z^2 \tau_c^4 + 15\pi_c \tau_c^6 - 3\pi_c^2 \tau_c^6 \\
\mathbb{V}[x_{ij}^4] &= 96\sigma_z^8 + 384\pi_c \sigma_z^6 \tau_c^2 + 612\pi_c \sigma_z^4 \tau_c^4 - 36\pi_c^2 \sigma_z^4 \tau_c^4 + 420\pi_c \sigma_z^2 \tau_c^6 - 36\pi_c^2 \sigma_z^2 \tau_c^6 + 105\pi_c \tau_c^8 - 9\pi_c^2 \tau_c^8.
\end{aligned}$$

Equation (15) yields the following variance:

$$\frac{1}{3} (8\sigma_z^8 + 32\pi_c \sigma_z^6 \tau_c^2 + 24(3 - \pi_c) \pi_c \sigma_z^4 \tau_c^4 + 16\pi_c (5 - 6\pi_c + 3\pi_c^2) \sigma_z^2 \tau_c^6 + \pi_c (35 - 63\pi_c + 48\pi_c^2 - 12\pi_c^3) \tau_c^8).$$

This can be simplified slightly to:

$$\frac{1}{3} \left( 8(\pi_c \tau_c^2 + \sigma_z^2)^4 + \pi_c (1 - \pi_c) (72\tau_c^4 \sigma_z^4 + 16(5 - \pi_c) \tau_c^6 \sigma_z^2 + (35 - 28\pi_c + 20) \tau_c^8) \right)$$

Now combining this with consistency of  $\hat{\sigma}_c^2 + \hat{\sigma}_z^2$  for  $\sigma_c^2 + \sigma_z^2 = \pi_c \tau_c^2 + \sigma_z^2$  as  $n, p \rightarrow \infty$  that follows directly from from Equations (17) and (13), we have:

$$\begin{aligned}
&\sqrt{np} \left( \frac{\hat{\sigma}_c^4}{\sqrt{\frac{8}{3}} (\hat{\sigma}_c^2 + \hat{\sigma}_z^2)^2} - \frac{\pi_c (1 - \pi_c) \tau_c^4}{\sqrt{\frac{8}{3}} (\pi_c \tau_c^2 + \sigma_z^2)^2} \right) \xrightarrow{d} \\
&N \left( 0, 1 + \frac{\pi_c (1 - \pi_c) (72\tau_c^4 \sigma_z^4 + 16(5 - \pi_c) \tau_c^6 \sigma_z^2 + (35 - 28\pi_c + 20) \tau_c^8)}{8(\pi_c \tau_c^2 + \sigma_z^2)^4} \right).
\end{aligned}$$

We can simplify the mean and variance and write them as a function of  $\phi^2 = \tau_c^2 / \sigma_z^2$ :

$$\begin{aligned}
&\frac{\pi_c (1 - \pi_c) \tau_c^4}{\sqrt{\frac{8}{3}} (\pi_c \tau_c^2 + \sigma_z^2)^2} = \frac{\pi_c (1 - \pi_c) \phi^4}{\sqrt{\frac{8}{3}} (\pi_c \phi^2 + 1)^2} \\
&1 + \frac{\pi_c (1 - \pi_c) (72\tau_c^4 \sigma_z^4 + 16(5 - \pi_c) \tau_c^6 \sigma_z^2 + (35 - 28\pi_c + 20) \tau_c^8)}{8(\pi_c \tau_c^2 + \sigma_z^2)^4} = \\
&1 + \frac{\pi_c (1 - \pi_c) (72\phi^4 + 16(5 - \pi_c) \phi^6 + (35 - 28\pi_c + 20) \phi^8)}{8(\pi_c \phi^2 + 1)^4}
\end{aligned}$$

It follows that:

$$\Pr \left( \sqrt{np} \left( \frac{\hat{\sigma}_c^4}{\sqrt{\frac{8}{3}} (\hat{\sigma}_c^2 + \hat{\sigma}_z^2)^2} \right) > z_{1-\alpha} \right) \rightarrow 1 - \Phi \left( \frac{z_{1-\alpha} - \sqrt{\frac{3np}{8}} \frac{\pi_c(1-\pi_c)\phi^4}{(\pi_c\phi^2+1)^2}}{\sqrt{1 + \frac{\pi_c(1-\pi_c)(72\phi^4+16(5-\pi_c)\phi^6+(35-28\pi_c+20\pi_c^2)\phi^8)}{8(\pi_c\phi^2+1)^4}}}} \right)$$

as  $n, p \rightarrow \infty$ .

**Proof of Proposition 3.4:** First, we consider the case where  $p \rightarrow \infty$  with  $n$  fixed. The case where  $p \rightarrow \infty$  with  $n$  fixed is analogous. In the proof of Proposition 2.2, we show that  $\bar{r}^{(2)} \xrightarrow{p} \left(\frac{n-1}{n}\right) (\sigma_c^2 + \sigma_z^2)$  and  $\bar{r}^{(4)} \xrightarrow{p} \kappa \left(\frac{(n-1)(n^2-3n+3)}{n^3}\right) \sigma_c^4 + 3 \left(\frac{n-1}{n}\right)^2 (\sigma_c^2 + \sigma_z^2)^2$  as  $p \rightarrow \infty$  with  $n$  fixed. Applying the continuous mapping theorem yields  $\hat{\sigma}_c^2 \xrightarrow{p} \sqrt{\kappa/3} \sigma_c^2$  and  $\hat{\sigma}_z^2 \xrightarrow{p} \sigma_z^2 + \left(1 - \sqrt{\kappa/3}\right) \sigma_z^2$  as  $p \rightarrow \infty$  with  $n$  fixed.

Now we consider the case where  $n, p \rightarrow \infty$ . In the proof of Proposition 2.2, we show that  $\bar{r}^{(2)} \xrightarrow{p} \sigma_c^2 + \sigma_z^2$  and  $\bar{r}^{(4)} \xrightarrow{p} \kappa \sigma_c^4 + 3 (\sigma_c^2 + \sigma_z^2)^2$  as  $n, p \rightarrow \infty$ . Applying the continuous mapping theorem yields  $\hat{\sigma}_c^2 \xrightarrow{p} \sqrt{\kappa/3} \sigma_c^2$  and  $\hat{\sigma}_z^2 \xrightarrow{p} \sigma_z^2 + \left(1 - \sqrt{\kappa/3}\right) \sigma_z^2$  as  $n, p \rightarrow \infty$ .

**Proof of Proposition 4.1:** We prove Proposition 4.1 by computing  $\mathbb{E} \left[ (\bar{r}^{(2)})^2 \right]$  and  $\mathbb{E} [\bar{r}^{(4)}]$ . Starting with  $\mathbb{E} \left[ (\bar{r}^{(2)})^2 \right]$ , we again use the result of Konig et al. (1992) given by Equation (9), which yields:

$$\begin{aligned} \mathbb{E} \left[ (\bar{r}^{(2)})^2 \right] &= \mathbb{E} \left[ \left( \text{vec}(\mathbf{C} + \mathbf{Z})' \mathbf{H}^{(K)} \text{vec}(\mathbf{C} + \mathbf{Z}) \right)^2 \right] \\ &= \alpha_K \left( \kappa \sigma_c^4 + 3 (\sigma_c^2 + \sigma_z^2)^2 \right) + (\theta_K + 2\beta_K) (\sigma_c^2 + \sigma_z^2)^2, \end{aligned} \quad (17)$$

where  $\alpha_K = \sum_{j=1}^p \left( h_{jj}^{(K)} \right)^2$ ,  $\beta_K = \sum_{j=1}^p \sum_{l=1, l \neq j}^p h_{jj}^{(K)} h_{ll}^{(K)}$  and  $\theta_K = \sum_{j=1}^p \sum_{l=1, l \neq j}^p \left( h_{jl}^{(K)} \right)^2$  and  $\kappa$  is the excess kurtosis for the distribution of the elements of  $\mathbf{C}$ , which in this case where elements of  $\mathbf{C}$  are assumed to be Laplace distributed is  $\kappa = 3$ .

From the definition of  $\mathbf{H}^{(K)}$  it is straightforward to see that all of the diagonal elements of  $\mathbf{H}^{(K)}$  are identical and equal to  $\prod_{k=1}^K \frac{p_k-1}{p_k}$ . Therefore, it is straightforward to compute  $\alpha_K$  and  $\theta_K$ .

$$\begin{aligned}
\alpha_K &= \prod_{k=1}^K p_K \left( \frac{p_K - 1}{p_K} \right)^2 \\
&= \prod_{k=1}^K \frac{(p_k - 1)^2}{p_k} \\
\theta_K &= \prod_{k=1}^K p_k \left( \prod_{k=1}^K p_k - 1 \right) \prod_{k=1}^K \left( \frac{p_k - 1}{p_k} \right)^2 \\
&= \left( \prod_{k=1}^K p_k - 1 \right) \prod_{k=1}^K \frac{(p_k - 1)^2}{p_k} \\
&= \prod_{k=1}^K (p_k - 1)^2 - \prod_{k=1}^K \frac{(p_k - 1)^2}{p_k}
\end{aligned}$$

Computing  $\beta_K$  is a little bit trickier, however we can rewrite  $\mathbf{H}^{(K)} = \mathbf{H}_K \otimes \mathbf{H}^{(K-1)}$  and obtain a formula for  $\beta_K$  as a function of  $\alpha_{K-1}$  and  $\beta_{K-1}$ :

$$\begin{aligned}
\beta_K &= p_K (p_K - 1) \left( \frac{1}{p_K^2} \right) (\alpha_{K-1} + \beta_{K-1}) + p_K \left( \frac{p_K - 1}{p_K} \right)^2 \beta_{K-1} \\
&= \left( \frac{p_K - 1}{p_K} \right) (\alpha_{K-1} + \beta_{K-1}) + \frac{(p_K - 1)^2}{p_K} \beta_{K-1} \\
&= \left( \frac{p_K - 1}{p_K} \right) \alpha_{K-1} + (p_K - 1) \beta_{K-1}
\end{aligned}$$

To eliminate  $\beta_{K-1}$  from this expression, we note that  $\alpha_K + \beta_K = \text{tr}(\mathbf{H}^{(K)} \mathbf{H}^{(K)}) = \text{tr}(\mathbf{H}^{(K)}) = \left( \prod_{k=1}^K p_k \right) \left( \prod_{k=1}^K \frac{p_k - 1}{p_k} \right) = \prod_{k=1}^K (p_k - 1)$ . Adding and subtracting  $(p_K - 1) \alpha_{K-1}$  yields:

$$\begin{aligned}
\beta_K &= \left( \frac{p_K - 1}{p_K} - (p_K - 1) \right) \alpha_{K-1} + (p_K - 1) (\alpha_{K-1} + \beta_{K-1}) \\
&= \left( \frac{p_K - 1}{p_K} - (p_K - 1) \right) \alpha_{K-1} + (p_K - 1) \prod_{k=1}^{K-1} (p_k - 1) \\
&= \left( \frac{2p_K - 1 - p_K^2}{p_K} \right) \prod_{k=1}^{K-1} \frac{(p_k - 1)^2}{p_k} + \prod_{k=1}^K (p_k - 1) \\
&= - \underbrace{\prod_{k=1}^K \frac{(p_k - 1)^2}{p_k}}_{\alpha_K} + \prod_{k=1}^K (p_k - 1).
\end{aligned}$$

Plugging these terms in to Equation (13) yields:

$$\begin{aligned}
\mathbb{E} \left[ (\bar{r}^{(2)})^2 \right] &= \prod_{k=1}^K \frac{(p_k - 1)^2}{p_k} \left( \kappa \sigma_c^4 + 3 (\sigma_c^2 + \sigma_z^2)^2 \right) + \\
&\quad \left( \prod_{k=1}^K (p_k - 1)^2 - \prod_{k=1}^K \frac{(p_k - 1)^2}{p_k} + \right. \\
&\quad \left. 2 \left( - \prod_{k=1}^K \frac{(p_k - 1)^2}{p_k} + \prod_{k=1}^K (p_k - 1) \right) \right) (\sigma_c^2 + \sigma_z^2)^2 \\
&= \prod_{k=1}^K \frac{(p_k - 1)^2}{p_k} \kappa \sigma_c^4 + \left( \prod_{k=1}^K (p_k - 1)^2 + \prod_{k=1}^K 2 (p_k - 1) \right) (\sigma_c^2 + \sigma_z^2)^2.
\end{aligned} \tag{18}$$

Now we need to compute  $\mathbb{E} [\bar{r}^{(4)}]$ . We start by rewriting the expression,

$$\begin{aligned}
\mathbb{E} [\bar{r}^{(4)}] &= \frac{1}{\prod_{k=1}^K p_k} \sum_{k=1}^K \sum_{i_k=1}^{p_k} \left( \text{vec}(\mathbf{R}) \mathbf{Q}^{(i_1, \dots, i_K)} \text{vec}(\mathbf{R}) \right)^2 \\
&= \frac{1}{\prod_{k=1}^K p_k} \sum_{i_1=1}^{p_1} \dots \sum_{i_K=1}^{p_K} \left( \text{vec}(\mathbf{C} + \mathbf{Z})' \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \text{vec}(\mathbf{C} + \mathbf{Z}) \right)^2
\end{aligned}$$

where  $\mathbf{Q}^{(i_1 \dots i_K)} = (\mathbf{e}_{i_K} \mathbf{e}_{i_K}' ) \otimes \dots \otimes (\mathbf{e}_{i_1} \mathbf{e}_{i_1}')$ , where  $\mathbf{e}_{i_k}$  is a  $p_k \times 1$  vector with entry  $i_k$  equal to 1 and all other entries equal to zero.

Again, we can use Equation (9):

$$\begin{aligned}\mathbb{E} [\bar{r}^{(4)}] &= \frac{1}{\prod_{k=1}^K p_k} \sum_{i_1=1}^{p_1} \cdots \sum_{i_K=1}^{p_K} \left( \sum_{j=1}^{\prod_{k=1}^K p_k} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jj}^2 \left( \kappa \sigma_c^4 + 3 (\sigma_c^2 + \sigma_z^2)^2 \right) + \right. \\ &\quad \sum_{j=1}^{\prod_{k=1}^K p_k} \sum_{l=1, l \neq j}^{\prod_{k=1}^K p_k} \left( \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jj} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{ll} + \right. \\ &\quad \left. \left. 2 \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jl}^2 \right) (\sigma_c^2 + \sigma_z^2)^2 \right).\end{aligned}$$

First, to simplify the problem, we show that

$$\sum_{j=1}^{\prod_{k=1}^K p_k} \sum_{l=1, l \neq j}^{\prod_{k=1}^K p_k} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jj} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{ll} = \sum_{j=1}^{\prod_{k=1}^K p_k} \sum_{l=1, l \neq j}^{\prod_{k=1}^K p_k} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jl}^2.$$

First, we compute the following:

$$\begin{aligned}\sum_{j=1}^{\prod_{k=1}^K p_k} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jj}^2 &+ \sum_{j=1}^{\prod_{k=1}^K p_k} \sum_{l=1, l \neq j}^{\prod_{k=1}^K p_k} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jj} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{ll} = \\ &\text{tr} \left( \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right) \otimes \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right) \right) = \\ &\text{tr} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)^2 = \\ &\text{tr} \left( \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)^2 = \\ &\prod_{k=1}^K \text{tr} \left( \mathbf{e}_{ik} \mathbf{e}_{ik}' \left( \mathbf{I}_{p_k} - \mathbf{1}_{p_k} \mathbf{1}_{p_k}' / p_k \right) \right)^2 = \\ &\prod_{k=1}^K \text{tr} \left( \left( \mathbf{e}_{ik}' - \mathbf{1}_{p_k}' / p_k \right) \mathbf{e}_{ik} \right)^2 = \\ &\prod_{k=1}^K \left( \frac{p_k - 1}{p_k} \right)^2.\end{aligned}$$

Now we compute:

$$\begin{aligned}
& \sum_{j=1}^{\prod_{k=1}^K p_k} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jj}^2 + \sum_{j=1}^{\prod_{k=1}^K p_k} \sum_{l=1, l \neq j}^{\prod_{k=1}^K p_k} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jl}^2 = \\
& \text{tr} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right) = \\
& \text{tr} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \right) = \\
& \prod_{k=1}^K \text{tr} \left( (\mathbf{I}_{p_k} - \mathbf{1}_{p_k} \mathbf{1}'_{p_k} / p_k) \mathbf{e}_{ik} \mathbf{e}'_{ik} (\mathbf{I}_{p_k} - \mathbf{1}_{p_k} \mathbf{1}'_{p_k} / p_k) \mathbf{e}_{ik} \mathbf{e}'_{ik} \right) = \\
& \prod_{k=1}^K \text{tr} \left( \mathbf{e}'_{ik} (\mathbf{e}_{ik} - \mathbf{1}_{p_k} / p_k) (\mathbf{e}'_{ik} - \mathbf{1}'_{p_k} / p_k) \mathbf{e}_{ik} \right) = \\
& \prod_{k=1}^K \left( \frac{p_k - 1}{p_k} \right)^2.
\end{aligned}$$

It follows that

$$\sum_{j=1}^{\prod_{k=1}^K p_k} \sum_{l=1, l \neq j}^{\prod_{k=1}^K p_k} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jj} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{ll} = \sum_{j=1}^{\prod_{k=1}^K p_k} \sum_{l=1, l \neq j}^{\prod_{k=1}^K p_k} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jl}^2.$$

Now, to determine if  $\sum_{j=1}^{\prod_{k=1}^K p_k} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jj}^2$  and  $\sum_{j=1}^{\prod_{k=1}^K p_k} \sum_{l=1, l \neq j}^{\prod_{k=1}^K p_k} \left( \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} \right)_{jl}^2$  depend on the values of  $i_1, \dots, i_K$ , we examine the structure of each  $\mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)}$ .

$$\begin{aligned}
\mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)} &= (\mathbf{H}_K \otimes \dots \otimes \mathbf{H}_1) \left( (\mathbf{e}_{i_K} \mathbf{e}'_{i_K}) \otimes \dots \otimes (\mathbf{e}_{i_1} \mathbf{e}'_{i_1}) \right) (\mathbf{H}_K \otimes \dots \otimes \mathbf{H}_1) \\
&= (\mathbf{H}_K \mathbf{e}_{i_K} \mathbf{e}'_{i_K} \mathbf{H}_K) \otimes \dots \otimes (\mathbf{H}_1 \mathbf{e}_{i_1} \mathbf{e}'_{i_1} \mathbf{H}_1).
\end{aligned}$$

Each matrix  $\mathbf{H}_k \mathbf{e}_{i_k} \mathbf{e}'_{i_k} \mathbf{H}_k$  is formed by multiplying the  $i_k$ -th elements of each row of  $\mathbf{H}_k$  by the  $i_k$ -th elements of each column of  $\mathbf{H}_k$ , yielding a matrix with:

- one diagonal element equal to  $\left( \frac{p_k - 1}{p_k} \right)^2$ ;
- $p_k - 1$  diagonal elements equal to  $\frac{1}{p_k^2}$ ;
- $2(p_k - 1)$  off-diagonal elements equal to  $-\left( \frac{p_k - 1}{p_k^2} \right)$ ;
- $p_k(p_k - 1) - 2(p_k - 1)$  elements equal to  $\frac{1}{p_k^2}$ .

Accordingly, the specific  $i_k$  used to construct the matrix  $\mathbf{H}_k \mathbf{e}_{i_k} \mathbf{e}'_{i_k} \mathbf{H}_k$  determines the relative locations of the diagonal elements and the relative locations of the off-diagonal elements, but not the composition and values of the off-diagonal elements and diagonal

elements. As a result, the sum of squared diagonal elements, the sum of pairwise product of different diagonal elements and the sum of squared off-diagonal elements for each submatrix will *not* depend on the specific  $\mathbf{e}_{i_k}$  used to construct the matrix.

Likewise, because the diagonal elements of  $\mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)}$  are given by the products of diagonal elements of the components,  $\mathbf{H}_k \mathbf{e}_{i_k} \mathbf{e}_{i_k}' \mathbf{H}_k$ , and the off-diagonal elements of  $\mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)}$  are given by the products of diagonal and off-diagonal or off-diagonal alone elements of the components,  $\mathbf{H}_k \mathbf{e}_{i_k} \mathbf{e}_{i_k}' \mathbf{H}_k$ , the sum of squared diagonal elements and the sum of squared off diagonal elements of  $\mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)}$  will not depend on the values of  $i_1, \dots, i_K$ .

Defining  $\mathbf{S}^{(K)} = \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)}$  and setting  $\alpha_K = \sum_{j=1}^p \left( s_{jj}^{(K)} \right)^2$ ,  $\theta_K = \sum_{j=1}^p \sum_{l=1, l \neq j}^p s_{jj}^{(K)} s_{ll}^{(K)}$  and  $\beta_K = \sum_{j=1}^p \sum_{l=1, k \neq j}^p \left( s_{jl}^{(K)} \right)^2$ , applying Equation (4.1) from Konig et al. (1992) yields: For reasons discussed below, we drop the  $(i_1, \dots, i_K)$  superscript on  $\mathbf{S}$  because the sums of squared entries of  $\mathbf{S}$  and sums of products of diagonal entries of  $\mathbf{S}$  that we are interested in computing do not depend on them and it simplifies notation. We explain why this is the case later on.

Now we simplify the notation. Let  $\mathbf{S}^{(K)} = \mathbf{H}^{(K)} \mathbf{Q}^{(i_1, \dots, i_K)} \mathbf{H}^{(K)}$ , dropping the  $i_1, \dots, i_K$  indices because the functions of  $\mathbf{S}^{(K)}$  we are interested in do not depend on them. Define  $\alpha_K = \sum_{j=1}^{\prod_{k=1}^K p_k} \left( s_{jj}^{(K)} \right)^2$  and  $\beta_K = \sum_{j=1}^{\prod_{k=1}^K p_k} \sum_{l=1, l \neq j}^{\prod_{k=1}^K p_k} \left( s_{jl}^{(K)} \right)^2$ . Note that we can simplify our expression for  $\mathbb{E} [\bar{r}^{(4)}]$  from earlier as follows:

$$\mathbb{E} [\bar{r}^{(4)}] = \alpha_K \left( \kappa \sigma_c^4 + 3 (\sigma_c^2 + \sigma_z^2)^2 \right) + 3 \beta_K (\sigma_c^2 + \sigma_z^2)^2. \quad (19)$$

To help us compute  $\alpha_K$  and  $\beta_K$ , we can expand and rewrite  $\mathbf{S}^{(K)}$  as follows to obtain a recursive formula for  $\mathbf{S}^{(K)}$  given  $\mathbf{S}^{(K-1)}$ , which will :

$$\begin{aligned} \mathbf{S}^{(K)} &= (\mathbf{H}_K \otimes \dots \otimes \mathbf{H}_1) \left( (\mathbf{e}_{i_K} \mathbf{e}_{i_K}') \otimes \dots \otimes (\mathbf{e}_{i_1} \mathbf{e}_{i_1}') \right) (\mathbf{H}_K \otimes \dots \otimes \mathbf{H}_1) \\ &= (\mathbf{H}_K \mathbf{e}_{i_1} \mathbf{e}_{i_1}' \mathbf{H}_K) \otimes \left( (\mathbf{H}_{K-1} \otimes \dots \otimes \mathbf{H}_1) \left( (\mathbf{e}_{i_{K-1}} \mathbf{e}_{i_{K-1}}') \otimes \dots \otimes (\mathbf{e}_{i_1} \mathbf{e}_{i_1}') \right) (\mathbf{H}_{K-1} \otimes \dots \otimes \mathbf{H}_1) \right) \\ &= (\mathbf{H}_K \mathbf{e}_{i_1} \mathbf{e}_{i_1}' \mathbf{H}_K) \otimes \mathbf{S}^{(K-1)}. \end{aligned}$$

We can compute  $\alpha_K$  recursively from this formula; the sum of squared diagonal elements of  $\mathbf{S}^{(K)}$  will equal the sum of each diagonal element of  $\mathbf{H}_K \mathbf{e}_{i_1} \mathbf{e}_{i_1}' \mathbf{H}_K$  squared multiplied by



the sum of squared diagonal elements of  $\mathbf{S}^{(K-1)}$ ,  $\alpha_{K-1}$ :

$$\begin{aligned}
\alpha_K &= \left( \frac{(p_K - 1)^4}{p_K^4} + \frac{(p_K - 1)}{p_K^4} \right) \alpha_{K-1} \\
&= \frac{(p_K - 1)}{p_K^4} (p_K^3 - 3p_K^2 + 3p_K) \alpha_{K-1} \\
&= \frac{(p_K - 1)}{p_K^3} (p_K^2 - 3p_K + 3) \alpha_{K-1} \\
&= \prod_{k=1}^K \frac{(p_k - 1)(p_k^2 - 3p_k + 3)}{p_k^3}
\end{aligned}$$

Now we can compute  $\beta_K$  in a similar way. The sum of squared off-diagonal elements of  $\mathbf{S}^{(K)}$  will be given by the sum of each diagonal element of  $\mathbf{H}_K \mathbf{e}_{i_1} \mathbf{e}_{i_1}' \mathbf{H}_K$  squared multiplied by the sum of squared off diagonal elements of  $\mathbf{S}^{(K-1)}$ ,  $\beta_{K-1}$ , plus the sum of each off-diagonal element of  $\mathbf{H}_K \mathbf{e}_{i_1} \mathbf{e}_{i_1}' \mathbf{H}_K$  squared multiplied by the sum of squared elements of  $\mathbf{S}^{(K-1)}$ ,  $\alpha_{K-1} + \beta_{K-1}$ .

$$\begin{aligned}
\beta_K &= \left( \left( \frac{p_K - 1}{p_K} \right)^4 + \frac{(p_K - 1)}{p_K^4} \right) \beta_{K-1} + \\
&\quad \left( \frac{2(p_K - 1)^3}{p_K^4} + \frac{p_K(p_K - 1) - 2(p_K - 1)}{p_K^4} \right) (\alpha_{K-1} + \beta_{K-1}) \\
&= \left( \frac{p_K - 1}{p_K^4} \right) ((p_K - 1)^3 + 1 + 2(p_K - 1)^2 + p_K - 2) \beta_{K-1} + \\
&\quad \left( \frac{p_K - 1}{p_K^4} \right) (2(p_K - 1)^2 + p_K - 2) \alpha_{K-1} \\
&= \left( \frac{p_K - 1}{p_K^4} \right) (p_K^3 - 3p_K^2 + 3p_K + 1 + 2p_K^2 - 4p_K + 2 + p_K - 2) \beta_{K-1} + \\
&\quad \left( \frac{p_K - 1}{p_K^4} \right) (2p_K^2 - 4p_K + 2 + p_K - 2) \alpha_{K-1} \\
&= \left( \frac{p_K - 1}{p_K^3} \right) ((p_K^2 - p_K)(\alpha_{K-1} + \beta_{K-1}) + (-p_K^2 + 3p_K - 3) \alpha_{K-1})
\end{aligned}$$

Noting that  $\alpha_{K-1} + \beta_{K-1} = \text{tr}(\mathbf{S}^{(K-1)} \mathbf{S}^{(K-1)})$ , which we computed earlier as equal to

$\prod_{k=1}^{K-1} \left( \frac{p_k-1}{p_k} \right)^2$ , we can eliminate  $\beta_{K-1}$  from this expression, yielding:

$$\begin{aligned}\beta_K &= \left( \frac{p_K-1}{p_K} \right)^2 \prod_{k=1}^{K-1} \left( \frac{p_k-1}{p_k} \right)^2 - \alpha_K \\ &= \prod_{k=1}^K \left( \frac{p_k-1}{p_k} \right)^2 - \alpha_K \\ &= \prod_{k=1}^K \left( \frac{p_k-1}{p_k} \right)^2 - \prod_{k=1}^K \frac{(p_k-1)(p_k^2-3p_k+3)}{p_k^3}.\end{aligned}$$

Substituting these quantities into Equation (19) yields:

$$\mathbb{E} [\bar{r}^{(4)}] = \prod_{k=1}^K \frac{(p_k-1)(p_k^2-3p_k+3)}{p_k^3} (\kappa \sigma_c^4) + \prod_{k=1}^K \left( \frac{p_k-1}{p_k} \right)^2 3 (\sigma_c^2 + \sigma_z^2)^2. \quad (20)$$

Proposition 4.1 follows directly from Equations (18) and (20) and plugging in  $\kappa = 3$ .

**Proof of Proposition 4.2:** First, we consider the scenario where a single  $p_{k'} \rightarrow \infty$ , with all other  $p_k$ ,  $k \neq k'$  held fixed. Without loss of generality, we consider  $p_K \rightarrow \infty$  and assume  $p_k$ ,  $k < K$ , fixed. First, we show consistency of  $\bar{r}^{(2)}$ , which involves first computing  $\mathbb{E} [\bar{r}^{(2)}]$  and then showing that  $\mathbb{V} [\bar{r}^{(2)}] = O\left(\frac{1}{p_K}\right)$ .

We compute  $\mathbb{E} [\bar{r}^{(2)}]$  as follows, where  $\mathbf{H}_k = \mathbf{I}_{p_k} - \mathbf{1}_{p_k} \mathbf{1}_{p_k}' / p_k$  and  $\mathbf{H}^{(K)} = \mathbf{H}_K \otimes \cdots \otimes \mathbf{H}_1$ :

$$\begin{aligned}\mathbb{E} [\bar{r}^{(2)}] &= \frac{1}{\prod_{k=1}^K p_k} \mathbb{E} \left[ \text{vec}(\mathbf{Y})' \mathbf{H}^{(K)} \text{vec}(\mathbf{Y}) \right] \\ &= \frac{1}{\prod_{k=1}^K p_k} \mathbb{E} \left[ \text{vec}(\mathbf{C} + \mathbf{Z})' \mathbf{H}^{(K)} \text{vec}(\mathbf{C} + \mathbf{Z}) \right] \\ &= \frac{1}{\prod_{k=1}^K p_k} \text{tr} \left( \mathbf{H}^{(K)} \right) (\sigma_c^2 + \sigma_z^2) \\ &= \prod_{k=1}^K \left( \frac{p_k-1}{p_k} \right) (\sigma_c^2 + \sigma_z^2).\end{aligned} \quad (21)$$

Now we turn to computing  $\mathbb{V} [\bar{r}^{(2)}]$ , which we write as:

$$\begin{aligned}\mathbb{V} [\bar{r}^{(2)}] &= \mathbb{E} \left[ (\bar{r}^{(2)})^2 \right] - \mathbb{E} [\bar{r}^{(2)}]^2 \\ &= \frac{1}{\prod_{k=1}^K p_k^2} \mathbb{E} \left[ \left( \text{vec}(\mathbf{C} + \mathbf{Z})' \mathbf{H}^{(K)} \text{vec}(\mathbf{C} + \mathbf{Z}) \right)^2 \right] - \prod_{k=1}^K \left( \frac{p_k-1}{p_k} \right)^2 (\sigma_c^2 + \sigma_z^2)^2.\end{aligned} \quad (22)$$

Substituting the expression for  $\mathbb{E} [(\bar{r}^{(2)})^2]$  given by Equation (18) into the expression for

the variance of  $\bar{r}^{(2)}$  yields:

$$\begin{aligned}\mathbb{V}[\bar{r}^{(2)}] &= \prod_{k=1}^K \frac{(p_k - 1)^2}{p_k^3} \kappa \sigma_c^4 + \prod_{k=1}^K \frac{2(p_k - 1)}{p_k^2} (\sigma_c^2 + \sigma_z^2)^2 \\ &= O\left(\frac{1}{p_K}\right).\end{aligned}$$

As a result,  $\bar{r}^{(2)} \xrightarrow{p} \prod_{k=1}^{K-1} \left(\frac{p_k - 1}{p_k}\right) (\sigma_c^2 + \sigma_z^2)$  as  $p_K \rightarrow \infty$  and  $p_k, k < K$  fixed. More generally, for fixed  $k'$ ,  $\bar{r}^{(2)} \xrightarrow{p} \prod_{k=1, k \neq k'}^K \left(\frac{p_k - 1}{p_k}\right) (\sigma_c^2 + \sigma_z^2)$  as  $p_{k'} \rightarrow \infty$  and  $p_k, k \neq k'$  fixed.

Now we show consistency of  $\bar{r}^{(4)}$ , which involves using the expression for  $\mathbb{E}[\bar{r}^{(2)}]$  given by Equation (20) and showing that  $\mathbb{V}[\bar{r}^{(4)}] = O\left(\frac{1}{p_K}\right)$ . First, we drop the  $(K)$  superscript from  $\mathbf{H}^{(K)}$  and let  $p = \prod_{k=1}^K p_k$  for simplicity. We decompose  $\mathbf{H}$ :  $\mathbf{H} = \mathbf{F} - \frac{1}{p_K} \mathbf{G}$ , where  $\mathbf{F} = \mathbf{I}_{p_K} \otimes \mathbf{H}_{K-1} \otimes \cdots \otimes \mathbf{H}_1$  and  $\mathbf{G} = (\mathbf{1}_{p_K} \mathbf{1}_{p_K}') \otimes \mathbf{H}_{K-1} \otimes \cdots \otimes \mathbf{H}_1$ . Accordingly, we have  $\text{vec}(\mathbf{R}) = \left(\mathbf{F} - \frac{1}{p_K} \mathbf{G}\right) \text{vec}(\mathbf{Y})$ .

Let  $\mathbf{f}_j$  and  $\mathbf{g}_j$  refer to the  $j$ -th rows of  $\mathbf{F}$  and  $\mathbf{G}$ , accordingly. Then the  $j$ -th element of  $\text{vec}(\mathbf{R})$ ,  $r_j = \mathbf{f}_j' \text{vec}(\mathbf{Y}) - \mathbf{g}_j' \text{vec}(\mathbf{Y}) / p_K$ . The operation  $\mathbf{g}_j' \text{vec}(\mathbf{Y}) / p_K$  takes means of elements of  $\mathbf{Y}$  over the  $K$ -th mode, so  $\mathbf{g}_j' \text{vec}(\mathbf{Y}) / p_K = O_p\left(\frac{1}{\sqrt{p_K}}\right)$ . Then:

$$\begin{aligned}\mathbb{V}[\bar{r}^{(4)}] &= \mathbb{V}\left[\frac{1}{p} \sum_{j=1}^p r_j^4\right] \\ &= \mathbb{E}\left[\left(\frac{1}{p} \sum_{j=1}^p \left(\mathbf{f}_j' \text{vec}(\mathbf{Y}) + O_p\left(\frac{1}{\sqrt{p_K}}\right)\right)^4\right)^2\right] - \left(\mathbb{E}\left[\frac{1}{p} \sum_{j=1}^p r_j^4\right]\right)^2\end{aligned}\tag{23}$$

Expanding the first term, we have:

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{p} \sum_{j=1}^p \left(\mathbf{f}_j' \text{vec}(\mathbf{Y}) + O_p\left(\frac{1}{\sqrt{p_K}}\right)\right)^4\right)^2\right] &= \\ \frac{1}{p^2} \sum_{j=1}^p \mathbb{E}\left[\left(\mathbf{f}_j' \text{vec}(\mathbf{Y}) + O_p\left(\frac{1}{\sqrt{p_K}}\right)\right)^8\right] &+ \\ \frac{1}{p^2} \sum_{j=1}^p \sum_{j'=1, j' \neq j}^p \mathbb{E}\left[\left(\mathbf{f}_j' \text{vec}(\mathbf{Y}) + O_p\left(\frac{1}{\sqrt{p_K}}\right)\right)^4 \left(\mathbf{f}_{j'}' \text{vec}(\mathbf{Y}) + O_p\left(\frac{1}{\sqrt{p_K}}\right)\right)^4\right].\end{aligned}$$

Now let's consider the structure of each  $\mathbf{f}_j$ , recalling that  $\mathbf{F} = \mathbf{I}_{p_K} \otimes \mathbf{H}_{K-1} \otimes \cdots \otimes \mathbf{H}_1$ . Each  $\mathbf{f}_j$  has  $p$  entries, but only  $\prod_{k=1}^{K-1} p_k$  entries are nonzero. If  $\text{vec}(\mathbf{Y})_j$  belongs to the  $i_j$ -th level of the  $K$ -th mode of  $\mathbf{Y}$ , the nonzero entries correspond to elements of  $\text{vec}(\mathbf{Y})$  that also belong to the  $i_j$ -th level of the  $K$ -th mode of  $\mathbf{Y}$ . This means that each product,  $\mathbf{f}_j' \text{vec}(\mathbf{Y})$ ,

includes only the  $\prod_{k=1}^{K-1} p_k$  entries of  $\text{vec}(\mathbf{Y})$  that belong to the  $i_j$ -th level of the  $K$ -th mode of  $\mathbf{Y}$ .

Because the number of elements of  $\mathbf{Y}$  in each  $\mathbf{f}'_j \text{vec}(\mathbf{Y})$ , the entries of any  $\mathbf{f}_j$ , and the eighth moment of each element of  $\mathbf{Y}$  are finite and do not depend at all on  $p_K$  and are fixed when  $p_1, \dots, p_{K-1}$  are fixed, each  $\mathbb{E}[(\mathbf{f}'_j \text{vec}(\mathbf{Y}))^a]$  for  $a = 1, \dots, 8$  are finite and fixed for fixed  $p_1, \dots, p_{K-1}$ . As a result,  $\mathbb{E}[(\mathbf{f}'_j \text{vec}(\mathbf{Y}))^a] O_p\left(\frac{1}{\sqrt{p_K}}\right) = O_p\left(\frac{1}{\sqrt{p_K}}\right)$ . Additionally, by exchangeability of the elements of  $\mathbf{Y}$  and structure of the matrix  $\mathbf{F}$  (each row of  $\mathbf{F}$  contains the same elements in a different order),  $\mathbb{E}[(\mathbf{f}'_j \text{vec}(\mathbf{Y}))^a] = \mathbb{E}[(\mathbf{f}'_1 \text{vec}(\mathbf{Y}))^a]$  for all  $j$  and  $\frac{1}{p^2} \sum_{j=1}^p \mathbb{E}[(\mathbf{f}'_j \text{vec}(\mathbf{Y}))^8] = \frac{1}{p} \mathbb{E}[(\mathbf{f}'_1 \text{vec}(\mathbf{Y}))^8] = O\left(\frac{1}{p_1}\right)$ . Then we can simplify the first term as follows:

$$\begin{aligned} & \mathbb{E} \left[ \left( \frac{1}{p} \sum_{j=1}^p \left( \mathbf{f}'_j \text{vec}(\mathbf{Y}) + O_p\left(\frac{1}{\sqrt{p_K}}\right) \right)^4 \right)^2 \right] = \\ & \frac{1}{p^2} \sum_{j=1}^p \sum_{j'=1, j' \neq j}^p \mathbb{E} \left[ (\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4 (\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4 \right] + O_p\left(\frac{1}{\sqrt{p_K}}\right). \end{aligned}$$

Now, we consider how the  $\mathbf{f}'_j \text{vec}(\mathbf{Y})$  and  $\mathbf{f}'_{j'} \text{vec}(\mathbf{Y})$  relate when  $j \neq j'$ , specifically with respect to which elements of  $\mathbf{Y}$  are contained in each and if there is any overlap. We define  $i_j$  as the level of the  $K$ -th mode that the element  $\text{vec}(\mathbf{Y})_j$  belongs to. For a fixed  $j$ , let  $\mathcal{J}_j$  denote the set of  $\prod_{k=1}^{K-1} p_k - 1$  indices that correspond to other entries of  $\mathbf{Y}$  belonging to the  $i_j$ -th level of the  $K$ -th mode of  $\mathbf{Y}$ , besides  $\text{vec}(\mathbf{Y})_j$ . Then if  $j' \notin \mathcal{J}_j$ ,  $\mathbf{f}'_{j'} \text{vec}(\mathbf{Y})$  does not contain any of the same elements of  $\mathbf{Y}$  as  $\mathbf{f}'_j \text{vec}(\mathbf{Y})$ , so we can write:

$$\begin{aligned} & \mathbb{E} \left[ \left( \frac{1}{p} \sum_{j=1}^p \left( \mathbf{f}'_j \text{vec}(\mathbf{Y}) + O_p\left(\frac{1}{\sqrt{p_K}}\right) \right)^4 \right)^2 \right] = \\ & \frac{1}{p^2} \sum_{j=1}^p \sum_{j' \in \mathcal{J}_j} \mathbb{E} \left[ (\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4 (\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4 \right] + \frac{1}{p^2} \sum_{j=1}^p \sum_{j' \notin \mathcal{J}_j} \mathbb{E} \left[ (\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4 \right] \mathbb{E} \left[ (\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4 \right] + \\ & O_p\left(\frac{1}{\sqrt{p_K}}\right) = \\ & \frac{1}{p^2} \sum_{j=1}^p \sum_{j' \in \mathcal{J}_j} \mathbb{E} \left[ (\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4 (\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4 \right] + \frac{(p_K - 1) \prod_{k=1}^{K-1} p_k}{p} \mathbb{E} \left[ (\mathbf{f}'_1 \text{vec}(\mathbf{Y}))^4 \right]^2 + O_p\left(\frac{1}{\sqrt{p_K}}\right) \\ & \frac{1}{p^2} \sum_{j=1}^p \sum_{j' \in \mathcal{J}_j} \mathbb{E} \left[ (\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4 (\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4 \right] + \mathbb{E} \left[ (\mathbf{f}'_1 \text{vec}(\mathbf{Y}))^4 \right]^2 + O_p\left(\frac{1}{\sqrt{p_K}}\right) \end{aligned}$$

Note that by the same arguments, we can expand the second term of Equation (23)

similarly:

$$\left( \mathbb{E} \left[ \frac{1}{p} \sum_{j=1}^p r_j^4 \right] \right)^2 = (\mathbb{E} [\mathbf{f}'_1 \text{vec}(\mathbf{Y})])^2 + O\left(\frac{1}{p_K}\right).$$

Then we have

$$\mathbb{V} [\bar{r}^{(4)}] = \frac{1}{p^2} \sum_{j=1}^p \sum_{j' \in \mathcal{J}_j} \mathbb{E} \left[ (\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4 (\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4 \right] + O_p \left( \frac{1}{\sqrt{p_K}} \right).$$

The last step is to address the first term in the above expression. Recall that for fixed  $j$ ,  $(\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4$  for  $j' \in \mathcal{J}_j$  contains the same  $\prod_{k=1}^{K-1} p_k$  elements of  $\mathbf{Y}$  as  $(\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4$ , with coefficients that do not depend on  $p_K$ . Therefore,  $(\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4 (\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4$  is a polynomial in  $\prod_{k=1}^{K-1} p_k$  elements of  $\mathbf{Y}$  which does not depend on  $p_K$  at all, therefore  $\mathbb{E} \left[ (\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4 (\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4 \right]$  is fixed and finite. It may not be the case that  $\mathbb{E} \left[ (\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4 (\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4 \right]$  is equal for any  $j, j' \in \mathcal{J}_j$ , but a maximum, which again does not depend on  $p_K$  and is fixed for fixed  $p_1, \dots, p_{K-1}$  exists. Let  $C = \max_{j, j' \in \mathcal{J}_j} \mathbb{E} \left[ (\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4 (\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4 \right]$ . Then

$$\begin{aligned} \frac{1}{p^2} \sum_{j=1}^p \sum_{j' \in \mathcal{J}_j} \mathbb{E} \left[ (\mathbf{f}'_j \text{vec}(\mathbf{Y}))^4 (\mathbf{f}'_{j'} \text{vec}(\mathbf{Y}))^4 \right] &\leq \frac{p \left( \prod_{k=1}^{K-1} p_k - 1 \right)}{p^2} C \\ &= \left( \frac{1}{p_K} + \frac{1}{p} \right) C = O\left(\frac{1}{p_K}\right). \end{aligned}$$

Accordingly,  $\mathbb{V} [\bar{r}^{(4)}] = O_p \left( \frac{1}{\sqrt{p_K}} \right)$ . Then

$$\bar{r}^{(4)} \xrightarrow{p} \prod_{k=1}^{K-1} \frac{(p_k - 1)(p_k^2 - 3p_k + 3)}{p_k^3} (\kappa \sigma_c^4) + \prod_{k=1}^{K-1} \left( \frac{p_k - 1}{p_k} \right)^2 3(\sigma_c^2 + \sigma_z^2)^2$$

as  $p_K \rightarrow \infty$  when  $p_k, k < K$  fixed.

Combining this result with  $\bar{r}^{(2)} \xrightarrow{p} \prod_{k=1}^{K-1} \left( \frac{p_k - 1}{p_k} \right) (\sigma_c^2 + \sigma_z^2)$  as  $p_K \rightarrow \infty$  when  $p_k, k < K$  fixed, plugging in  $\kappa = 3$  and applying the continuous mapping theorem yields  $\hat{\sigma}_c^4 \xrightarrow{p} \sigma_c^4$ ,  $\hat{\sigma}_c^2 \xrightarrow{p} \sigma_c^2$ ,  $\hat{\lambda}_c \xrightarrow{p} \lambda_c$  and  $\hat{\sigma}_z^2 \xrightarrow{p} \sigma_z^2$ .

Now we consider the scenario where  $p_1, \dots, p_K \rightarrow \infty$ . It follows directly from Equations (21) and (22) that  $\bar{r}^{(2)} \xrightarrow{p} \sigma_c^2 + \sigma_z^2$  as  $p_1, \dots, p_K \rightarrow \infty$ . Next, we need to examine  $\bar{r}^{(4)}$ . For convenience, we also let  $p_k = \pi_k q$ , where  $0 < \pi_k < 1$  are fixed and  $\sum_{k=1}^K \pi_k = 1$ . We also let  $\pi = \prod_{k=1}^K \pi_k$ . Then the scenario  $p_1, \dots, p_K \rightarrow \infty$  is equivalent to  $q \rightarrow \infty$ . We again let  $x_{i_1 \dots i_K} = c_{i_1 \dots i_K} + z_{i_1 \dots i_K}$ . In this case, we do not give an explicit decomposition of elements of  $\mathbf{R}$  into elements of  $\mathbf{X}$ , rather we note that each element of  $\mathbf{R}$  is a sum of the corresponding element of  $\mathbf{X}$  and averages over terms of  $\mathbf{X}$  that belong to at least one of the same modes.

The averages with the largest variances are given by averages over the levels of just one mode, e.g.  $\bar{x}_{i_2 \dots i_K} = \frac{1}{\pi_1 q} \sum_{i_1=1}^{\pi_1 q} x_{i_1 \dots i_K} = O_p\left(\frac{1}{\sqrt{q}}\right)$ . Using the same logic as Equation (14) we can write:

$$\begin{aligned}\bar{r}^{(4)} &= \frac{1}{\pi q^K} \sum_{k=1}^K \sum_{i_k=1}^{\pi_k q} \left( x_{i_1 \dots i_K} + O_p\left(\frac{1}{\sqrt{q}}\right) \right)^4 \\ &= \bar{x}^{(4)} + O_p\left(\frac{1}{q}\right) \\ &= \kappa \sigma_c^4 + 3(\sigma_c^2 + \sigma_z^2)^2.\end{aligned}$$

It follows that  $\bar{r}^{(4)} \xrightarrow{p} \kappa \sigma_c^4 + 3(\sigma_c^2 + \sigma_z^2)^2$  as  $q \rightarrow \infty$ , i.e. as  $p_1, \dots, p_K \rightarrow \infty$ .

Plugging in the excess kurtosis,  $\kappa = 3$ , of Laplace distributed  $c_{ij}$ , combining  $\bar{r}^{(4)} \xrightarrow{p} 3\sigma_c^4 + 3(\sigma_c^2 + \sigma_z^2)^2$  as  $p_1, \dots, p_K \rightarrow \infty$  with  $\bar{r}^{(2)} \xrightarrow{p} \sigma_c^2 + \sigma_z^2$  as  $p_1, \dots, p_K \rightarrow \infty$  and applying the continuous mapping theorem yields  $\hat{\sigma}_c^4 \xrightarrow{p} \sigma_c^4$ ,  $\hat{\sigma}_c^2 \xrightarrow{p} \sigma_c^2$ ,  $\hat{\lambda}_c \xrightarrow{p} \lambda_c$  and  $\hat{\sigma}_z^2 \xrightarrow{p} \sigma_z^2$  as  $p_1, \dots, p_K \rightarrow \infty$ .

**Notes on Extending Propositions 3.1-3.3 to Tensor Data:** Extending Propositions 3.1-3.3 to tensor data amounts to showing

$$\sqrt{p}(\hat{\sigma}_c^4 - \kappa \sigma_c^4/3) \xrightarrow{d} \sqrt{p}(\bar{x}^{(4)}/3 - (\bar{x}^{(2)})^2 - \kappa \sigma_c^4/3),$$

where  $x_{i_1 \dots i_K} = c_{i_1 \dots i_K} + z_{i_1 \dots i_K}$ , as  $p_1, \dots, p_K \rightarrow \infty$ . We do not go into great detail showing this as we did in the matrix case because all the same logic used to prove  $\bar{r}^{(2)} \xrightarrow{d} \bar{x}^{(2)}$  and  $\bar{r}^{(4)} \xrightarrow{d} \bar{x}^{(4)}$  as  $n, p \rightarrow \infty$  in the matrix case can be applied.

For convenience, we also let  $p_k = \pi_k q$ , where  $0 < \pi_k < 1$  are fixed and  $\sum_{k=1}^K \pi_k = 1$ . We also let  $\pi = \prod_{k=1}^K \pi_k$ . Then the scenario  $p_1, \dots, p_K \rightarrow \infty$  is equivalent to  $q \rightarrow \infty$ . As in the previous proof, we do not give an explicit decomposition of elements of  $\mathbf{R}$  into elements of  $\mathbf{X}$ , rather we note that each element of  $\mathbf{R}$  is a sum of the corresponding element of  $\mathbf{X}$  and averages over terms of  $\mathbf{X}$  that belong to at least one of the same modes. The averages with the largest variances are given by averages over the levels of just one mode, e.g.  $\bar{x}_{i_2 \dots i_K} = \frac{1}{\pi_1 q} \sum_{i_1=1}^{\pi_1 q} x_{i_1 \dots i_K} = O_p\left(\frac{1}{\sqrt{q}}\right)$ , and the averages with the second largest variances are given by averages over the levels of two modes, e.g.  $\bar{x}_{i_3 \dots i_K} = \frac{1}{\pi_1 \pi_2 q^2} \sum_{i_1=1}^{\pi_1 q} \sum_{i_2=1}^{\pi_2 q} x_{i_1 i_2 \dots i_K} = O_p\left(\frac{1}{q}\right)$ . For ease of notation, we'll write the average of elements of  $\mathbf{X}$  over level  $i_k$  of the  $k$ -th mode as  $\bar{x}_{-i_k}$ . We can write:

$$r_{i_1 \dots i_K} = x_{i_1 \dots i_K} - \sum_{k=1}^K \bar{x}_{-i_k} + O_p\left(\frac{1}{q}\right).$$

For  $\bar{r}^{(2)}$  we have:

$$\begin{aligned}
\bar{r}^{(2)} &= \frac{1}{\pi q^K} \sum_{k=1}^K \sum_{i_k=1}^{\pi_k q} \left( x_{i_1 \dots i_K} - \sum_{k=1}^K \bar{x}_{-i_k} + O_p\left(\frac{1}{q}\right) \right)^2 \\
&= \frac{1}{\pi q^K} \sum_{k=1}^K \sum_{i_k=1}^{\pi_k q} \left( x_{i_1 \dots i_K} - \sum_{k=1}^K \bar{x}_{-i_k} \right)^2 + \left( x_{i_1 \dots i_K} - \sum_{k=1}^K \bar{x}_{-i_k} \right) O_p\left(\frac{1}{q}\right) + O_p\left(\frac{1}{q^2}\right) \\
&= \frac{1}{\pi q^K} \sum_{k=1}^K \sum_{i_k=1}^{\pi_k q} \left( x_{i_1 \dots i_K} - \sum_{k=1}^K \bar{x}_{-i_k} \right)^2 + O_p\left(\frac{1}{q^2}\right) \\
&= \frac{1}{\pi q^K} \sum_{k=1}^K \sum_{i_k=1}^{\pi_k q} x_{i_1 \dots i_K}^2 + \sum_{k=1}^K \bar{x}_{-i_k}^2 - 2 \sum_{k=1}^K x_{i_1 \dots i_K} \bar{x}_{-i_k} + 2 \sum_{k=1}^K \sum_{k'=1, k' \neq k}^K \bar{x}_{-i_k} \bar{x}_{-i'_k} + O_p\left(\frac{1}{q^2}\right) \\
&= \bar{x}^{(2)} + \frac{\mathbb{E}[x_{i_1 \dots i_K}^2]}{\pi q} + O_p\left(\frac{1}{\sqrt{q^3}}\right),
\end{aligned}$$

where the last line follows from using the same logic as the matrix case.

For  $\bar{r}^{(4)}$  we have:

$$\begin{aligned}
\bar{r}^{(4)} &= \frac{1}{\pi q^K} \sum_{k=1}^K \sum_{i_k=1}^{\pi_k q} \left( x_{i_1 \dots i_K} - \sum_{k=1}^K \bar{x}_{-i_k} + O_p\left(\frac{1}{q}\right) \right)^4 \\
&= \frac{1}{\pi q^K} \sum_{k=1}^K \sum_{i_k=1}^{\pi_k q} \left( x_{i_1 \dots i_K} - \sum_{k=1}^K \bar{x}_{-i_k} \right)^4 + \left( x_{i_1 \dots i_K} - \sum_{k=1}^K \bar{x}_{-i_k} \right)^3 O_p\left(\frac{1}{q}\right) + \\
&\quad \left( x_{i_1 \dots i_K} - \sum_{k=1}^K \bar{x}_{-i_k} \right)^2 O_p\left(\frac{1}{q^2}\right) + \left( x_{i_1 \dots i_K} - \sum_{k=1}^K \bar{x}_{-i_k} \right) O_p\left(\frac{1}{q^3}\right) + O_p\left(\frac{1}{q^4}\right) \\
&= \frac{1}{\pi q^K} \sum_{k=1}^K \sum_{i_k=1}^{\pi_k q} \left( x_{i_1 \dots i_K} - \sum_{k=1}^K \bar{x}_{-i_k} \right)^4 + O_p\left(\frac{1}{q^2}\right) \\
&= \frac{1}{\pi q^K} \sum_{k=1}^K \sum_{i_k=1}^{\pi_k q} x_{i_1 \dots i_K}^2 + \sum_{k=1}^K \bar{x}_{-i_k}^2 - 2 \sum_{k=1}^K x_{i_1 \dots i_K} \bar{x}_{-i_k} + 2 \sum_{k=1}^K \sum_{k'=1, k' \neq k}^K \bar{x}_{-i_k} \bar{x}_{-i'_k} + O_p\left(\frac{1}{q^2}\right) \\
&= \bar{x}^{(4)} + \frac{6\mathbb{E}[x_{i_1 \dots i_K}^2]}{\pi q} + O\left(\frac{1}{q^2}\right) + O_p\left(\frac{1}{\sqrt{q^3}}\right),
\end{aligned}$$

where again the last line follows from using the same logic as the matrix case.

Now we plug these expressions for  $\bar{r}^{(2)}$  and  $\bar{r}^{(4)}$  into our equation for  $\hat{\sigma}_c^4$ :

$$\begin{aligned}\hat{\sigma}_c^4 &= O(1) \left( \bar{r}^{(4)}/3 - (\bar{r}^{(2)})^2 \right) \\ &= O(1) \left( \bar{x}^{(4)}/3 - (\bar{x}^{(2)})^2 + O\left(\frac{1}{q^2}\right) + O_p\left(\frac{1}{\sqrt{q^3}}\right) \right).\end{aligned}$$

Recalling that  $\kappa$  is defined such that  $\mathbb{E}[x_{ij}^4]/3 - \mathbb{E}[x_{ij}^2]^2 = \kappa\sigma_c^4/3$ , we have

$$\sqrt{\pi q^K} (\hat{\sigma}_c^4 - \kappa\sigma_c^4/3) \xrightarrow{d} \sqrt{\pi q^K} \left( \bar{x}^{(4)}/3 - (\bar{x}^{(2)})^2 - \kappa\sigma_c^4/3 \right)$$

as  $q \rightarrow \infty$ , i.e.

$$\sqrt{p} (\hat{\sigma}_c^4 - \kappa\sigma_c^4/3) \xrightarrow{d} \sqrt{p} \left( \bar{x}^{(4)}/3 - (\bar{x}^{(2)})^2 - \kappa\sigma_c^4/3 \right).$$

as  $p_1, \dots, p_K \rightarrow \infty$ .

**Lower-Order Mean Parameter Variance Estimators for Three-Way Tensors:** First we define the following unpenalized OLS estimates:

$$\begin{aligned}\check{\mathbf{a}} &= (\mathbf{1}_{p_3}/p_3 \otimes \mathbf{1}_{p_2}/p_2 \otimes \mathbf{H}_{p_1}) \text{vec}(\mathbf{Y}) \\ \check{\mathbf{b}} &= (\mathbf{1}_{p_3}/p_3 \otimes \mathbf{H}_{p_2} \otimes \mathbf{1}_{p_1}/p_1) \text{vec}(\mathbf{Y}) \\ \check{\mathbf{d}} &= (\mathbf{H}_{p_3} \otimes \mathbf{1}_{p_2}/p_2 \otimes \mathbf{1}_{p_1}/p_1) \text{vec}(\mathbf{Y}) \\ \text{vec}(\check{\mathbf{E}}) &= (\mathbf{1}_{p_3}/p_3 \otimes \mathbf{H}_{p_2} \otimes \mathbf{H}_{p_1}) \text{vec}(\mathbf{Y}) \\ \text{vec}(\check{\mathbf{F}}) &= (\mathbf{H}_{p_3} \otimes \mathbf{1}_{p_2}/p_2 \otimes \mathbf{H}_{p_1}) \text{vec}(\mathbf{Y}) \\ \text{vec}(\check{\mathbf{G}}) &= (\mathbf{H}_{p_3} \otimes \mathbf{H}_{p_2} \otimes \mathbf{1}_{p_1}/p_1) \text{vec}(\mathbf{Y}).\end{aligned}$$

Now we define the following estimators of lower order-mean parameter variances:

$$\begin{aligned}\hat{\sigma}_a^2 &= \frac{p_1}{p_1 - 1} \left( \bar{\check{a}}^{(2)} - \frac{1}{p_2 - 1} \bar{\check{e}}^{(2)} - \frac{1}{p_3 - 1} \bar{\check{f}}^{(2)} + \frac{1}{(p_2 - 1)(p_3 - 1)} \bar{\check{r}}^{(2)} \right) \\ \hat{\sigma}_b^2 &= \frac{p_2}{p_2 - 1} \left( \bar{\check{b}}^{(2)} - \frac{1}{p_1 - 1} \bar{\check{e}}^{(2)} - \frac{1}{p_3 - 1} \bar{\check{g}}^{(2)} + \frac{1}{(p_1 - 1)(p_3 - 1)} \bar{\check{r}}^{(2)} \right) \\ \hat{\sigma}_d^2 &= \frac{p_3}{p_3 - 1} \left( \bar{\check{d}}^{(2)} - \frac{1}{p_1 - 1} \bar{\check{f}}^{(2)} - \frac{1}{p_2 - 1} \bar{\check{g}}^{(2)} + \frac{1}{(p_1 - 1)(p_2 - 1)} \bar{\check{r}}^{(2)} \right) \\ \hat{\sigma}_e^2 &= \frac{p_1 p_2}{(p_1 - 1)(p_2 - 1)} \left( \bar{\check{e}}^{(2)} - \frac{1}{p_3 - 1} \bar{\check{r}}^{(2)} \right) \\ \hat{\sigma}_f^2 &= \frac{p_1 p_3}{(p_1 - 1)(p_3 - 1)} \left( \bar{\check{f}}^{(2)} - \frac{1}{p_2 - 1} \bar{\check{r}}^{(2)} \right) \\ \hat{\sigma}_g^2 &= \frac{p_2 p_3}{(p_2 - 1)(p_3 - 1)} \left( \bar{\check{g}}^{(2)} - \frac{1}{p_1 - 1} \bar{\check{r}}^{(2)} \right).\end{aligned}$$



where  $\bar{\bar{a}}^{(2)} = \frac{1}{n} \sum_{i=1}^n \check{a}_i^2$ ,  $\bar{\bar{d}}^{(2)} = \frac{1}{np_2} \sum_{i=1}^n \sum_{j=1}^{p_2} \check{d}_{ij}^2$  and  $\bar{\bar{b}}^{(2)}$ ,  $\bar{\bar{c}}^{(2)}$ ,  $\bar{\bar{e}}^{(2)}$  and  $\bar{\bar{f}}^{(2)}$  are defined accordingly and  $\hat{\sigma}_a^2$  estimates  $\sigma_a^2 = 2/\lambda_a^2$ ,  $\sigma_b^2$ ,  $\sigma_c^2$ ,  $\sigma_d^2$ ,  $\sigma_e^2$  and  $\sigma_f^2$ . These estimators are unbiased and consistent under certain asymptotic scenarios; all of the estimators are consistent as  $p_1, \dots, p_K \rightarrow \infty$ , however only  $\hat{\sigma}_a^2$ ,  $\hat{\sigma}_e^2$  and  $\hat{\sigma}_f^2$  are consistent as  $p_1 \rightarrow \infty$  with  $p_2, p_3$  fixed.